

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
14. Juni 2001 (14.06.2001)

PCT

(10) Internationale Veröffentlichungsnummer  
**WO 01/42493 A2**

(51) Internationale Patentklassifikation<sup>7</sup>: **C12Q 1/68**

(21) Internationales Aktenzeichen: **PCT/DE00/04381**

(22) Internationales Anmeldedatum:  
6. Dezember 2000 (06.12.2000)

(25) Einreichungssprache: **Deutsch**

(26) Veröffentlichungssprache: **Deutsch**

(30) Angaben zur Priorität:  
199 59 691.3 6. Dezember 1999 (06.12.1999) **DE**

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von  
US): **EPIGENOMICS AG [DE/DE];** Kastanienallee 24,  
D-10435 Berlin (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): **OLEK, Alexander**  
[DE/DE]; Kyffhäuserstrasse 20, D-10781 Berlin (DE).  
**PIEPENBROCK, Christian [DE/DE];** Schwartzkopfs-  
trasse 7 B, D-10115 Berlin (DE).

(74) Anwalt: **SCHUBERT, Klemens;** Joachimstrasse 9,  
D-10119 Berlin (DE).

(81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,  
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,  
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,  
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,  
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Bestimmungsstaaten (regional): ARIPO-Patent (GH,  
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW),  
eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ,  
TM), europäisches Patent (AT, BE, CH, CY, DE, DK,  
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR),  
OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML,  
MR, NE, SN, TD, TG).

Veröffentlicht:

— Ohne internationalen Recherchenbericht und erneut zu  
veröffentlichen nach Erhalt des Berichts.

Zur Erklärung der Zweibuchstaben-Codes, und der anderen  
Abkürzungen wird auf die Erklärungen ("Guidance Notes on  
Codes and Abbreviations") am Anfang jeder regulären Ausgabe  
der PCT-Gazette verwiesen.

(54) Title: **METHOD FOR THE PARALLEL DETECTION OF THE DEGREE OF METHYLATION OF GENOMIC DNA**

(54) Bezeichnung: **VERFAHREN ZUR PARALLELEN DETEKTION DES METHYLIERUNGSZUSTANDES VON GENOMI-  
SCHER DNA**

(57) Abstract: The invention relates to a method for the parallel detection of the degree of methylation of genomic DNA wherein the following steps are performed: (a) chemical treatment at the 5' position of non-methylated cytosine bases converts said bases into uracil, thymidine or another base which exhibits hybridization behavior different to that of cytosine in a genomic DNA sample; (b) more than ten different fragments, each having less than 2000 base pairs in said chemically treated genomic DNA sample, are amplified simultaneously using synthetic oligonucleotides as a primer, whereby said primers each contain genomic sequences which are involved in gene regulation and/or transcribed and/or translated, such as those sequences which should be obtained after execution of steps (a); (c) the sequence contexts of all or a portion of the CpG dinucleotides or CpNpG trinucleotides contained in the amplified fragments are determined.

(57) Zusammenfassung: Beschrieben ist ein Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA bei dem man folgende Schritte ausführt: (a) in einer genomischen DNA Probe wandelt man durch chemische Behandlung an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base um; (b) aus dieser chemisch behandelten genomischen DNA amplifiziert man mehr als zehn unterschiedliche Fragmente, die jeweils weniger als 2000 Basenpaare lang sind, gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer, wobei diese Primer jeweils Sequenzen aus an der Genregulation beteiligten und/oder transkribierten und/oder translatierten genomischen Sequenzen enthalten, wie sie nach einer Behandlung gemäß Schritt (a) vorliegen würden; (c) man bestimmt den Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.

WO 01/42493 A2

## **Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA**

Die vorliegende Erfindung betrifft ein Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA.

Die nach den methodischen Entwicklungen der letzten Jahre in der Molekularbiologie gut studierten Beobachtungsebenen sind die Gene selbst, die Übersetzung dieser Gene in RNA und die daraus entstehenden Proteine. Wann im Laufe der Entwicklung eines Individuums welches Gen angeschaltet wird und wie Aktivieren und Inhibieren bestimmter Gene in bestimmten Zellen und Geweben gesteuert wird, ist mit hoher Wahrscheinlichkeit mit Ausmaß und Charakter der Methylierung der Gene bzw. des Genoms korrelierbar. Insofern ist die Annahme naheliegend, daß pathogene Zustände sich in einem veränderten Methylierungsmuster einzelner Gene oder des Genoms äußern.

Stand der Technik sind Verfahren, welche das Studium von Methylierungsmustern einzelner Gene gestatten. Jüngere Fortentwicklungen dieser Methode erlauben auch die Analyse kleinster Mengen an Ausgangsmaterial. Die vorliegende Erfindung beschreibt ein Verfahren zur parallelen Detektion des Methylierungszustandes genomischer DNA Proben, wobei ausgehend von einer Probe gleichzeitig zahlreiche verschiedene Fragmente aus an der Genregulation beteiligten oder/und transkribierten und/oder translatierten Sequenzen amplifiziert werden und anschließend der Sequenzkontext in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide untersucht wird.

5-Methylcytosin ist die häufigste kovalent modifizierte Base in der DNA eukaryotischer Zellen. Sie spielt beispielsweise eine Rolle in der Regulation der Transkription, genomischem Imprinting und in der Tumorgenese. Die Identifizierung von 5-Methylcytosin als Bestandteil genetischer Information ist daher von erheblichem Interesse. 5-Methylcytosin-Positionen können jedoch nicht durch Sequenzierung identifiziert werden, da 5-Methylcytosin das gleiche Basenpaarungsverhalten aufweist

wie Cytosin. Darüber hinaus geht bei einer PCR-Amplifikation die epigenetische Information, welche die 5-Methylcytosine tragen, vollständig verloren.

Die Modifikation der genomischen Base Cytosin zu 5'-Methylcytosin stellt den bis heute wichtigsten und best-untersuchten epigenetischen Parameter dar. Trotzdem gibt es bis heute zwar Methoden, umfassende Genotypen von Zellen und Individuen zu ermitteln, aber noch keine vergleichbaren Ansätze auch in großem Maße epigenotypische Information zu generieren und auszuwerten.

Es sind im Prinzip drei prinzipiell verschiedene Methoden bekannt, den 5-Methyl-Status eines Cytosins im Sequenzkontext zu bestimmen.

Die erste prinzipielle Methode beruht auf der Verwendung von Restriktionsendonukleasen (RE), welche „methylierungssensitiv“ sind. REs zeichnen sich dadurch aus, daß sie an einer bestimmten DNA-Sequenz, meist zwischen 4 und 8 Basen lang, einen Schnitt in die DNA einführen. Die Position solcher Schnitte kann dann durch Gelelektrophorese, Transfer auf eine Membran und Hybridisierung nachgewiesen werden. Methylierungssensitiv bedeutet, daß bestimmte Basen innerhalb der Erkennungssequenz unmethyliert vorliegen müssen, damit der Schnitt erfolgen kann. Das Bandenmuster nach einem Restriktionsschnitt und Gelelektrophorese ändert sich also je nach Methylierungsmuster der DNA. Allerdings befinden sich die wenigsten methylierbaren CpG innerhalb von Erkennungssequenzen von REs, können also nach dieser Methode nicht untersucht werden.

Die Empfindlichkeit dieser Methoden ist extrem niedrig (Bird, A.P., and Southern, E.M., J.Mol. Biol. 118, 27-47). Eine Variante kombiniert PCR mit dieser Methode, eine Amplifikation durch zwei auf beiden Seiten der Erkennungssequenz liegende Primer erfolgt nach einem Schnitt nur dann, wenn die Erkennungssequenz methyliert vorliegt. Die Empfindlichkeit steigt in diesem Fall auf theoretisch ein einziges Molekül der Zielsequenz, allerdings können mit hohem Aufwand nur einzelne Positionen untersucht werden (Shemer, R. et al., PNAS 93, 6371-6376). Wiederum ist Voraussetzung, daß sich die methylierbare Position innerhalb der Erkennungssequenz einer RE befindet.

Die zweite Variante beruht auf partieller chemischer Spaltung von Gesamt-DNA, nach dem Vorbild einer Maxam-Gilbert Sequenzierreaktion, Ligation von Adaptoren an die so generierten Enden, Amplifikation mit generischen Primern und Auftrennung auf einer Gelelektrophorese. Mit diesem Verfahren können definierte Bereiche bis zur Größe von weniger als tausend Basenpaaren untersucht werden. Das Verfahren ist allerdings so kompliziert und unzuverlässig, daß es praktisch nicht mehr verwendet wird (Ward, C. et al., J. Biol. Chem. 265, 3030-3033).

Eine relativ neue und die mittlerweile am häufigsten angewandte Methode zur Untersuchung von DNA auf 5-Methylcytosin beruht auf der spezifischen Reaktion von Bisulphit mit Cytosin, das nach anschließender alkalischer Hydrolyse in Uracil umgewandelt wird, welches in seinem Basen-Paarungsverhalten dem Thymidin entspricht. 5-Methylcytosin wird dagegen unter diesen Bedingungen nicht modifiziert. Damit wird die ursprüngliche DNA so umgewandelt, daß Methylcytosin, welches ursprünglich durch sein Hybridisierungsverhalten vom Cytosin nicht unterschieden werden kann, jetzt durch „normale“ molekularbiologische Techniken als einzig verbliebenes Cytosin beispielsweise durch Amplifikation und Hybridisierung oder Sequenzierung nachgewiesen werden kann. Alle diese Techniken beruhen auf Basenpaarung, welche jetzt voll ausgenutzt werden kann. Der Stand der Technik, was die Empfindlichkeit betrifft, wird durch ein Verfahren definiert, welches die zu untersuchende DNA in einer Agarose-Matrix einschließt, dadurch die Diffusion und Renaturierung der DNA (Bisulphit reagiert nur an einzelsträngiger DNA) verhindert und alle Fällungs- und Reinigungsschritte durch schnelle Dialyse ersetzt (Olek, A. et al., Nucl. Acids. Res. 24, 5064-5066). Mit dieser Methode können einzelne Zellen untersucht werden, was das Potential der Methode veranschaulicht. Allerdings werden bisher nur einzelne Regionen bis etwa 3000 Basenpaare Länge untersucht, eine globale Untersuchung von Zellen auf Tausende von möglichen Methylierungsereignissen ist nicht möglich. Allerdings kann auch dieses Verfahren keine sehr kleinen Fragmente aus geringen Probenmengen zuverlässig analysieren. Diese gehen trotz Diffusionsschutz durch die Matrix verloren.

Eine Übersicht über die weiteren bekannte Möglichkeiten, 5-Methylcytosine

nachzuweisen, kann auch dem folgenden Übersichtsartikel entnommen werden: Rein, T., DePamphilis, M. L., Zorbas, H., Nucleic Acids Res. 26, 2255 (1998).

Die Bisulphit-Technik wird bisher bis auf wenige Ausnahmen (z.B. Zeschnigk, M. et al., Eur. J. Hum. Gen. 5, 94-98; Kubota T. et al., Nat. Genet. 16, 16-17) nur in der Forschung angewendet. Immer aber werden kurze, spezifische Stücke eines bekannten Gens nach einer Bisulphit-Behandlung amplifiziert und entweder komplett sequenziert (Olek, A. and Walter, J., Nat. Genet. 17, 275-276) oder einzelne Cytosin-Positionen durch eine „Primer-Extension-Reaktion“ (Gonzalzo, M. L. and Jones, P.A., Nucl. Acids. Res. 25, 2529-2531) oder Enzymschnitt (Xiong, Z. and Laird, P.W., Nucl. Acids. Res. 25, 2532-2534) nachgewiesen. Zudem ist auch der Nachweis durch Hybridisierung beschrieben worden (Olek et al, WO9928498).

Gemeinsamkeiten zwischen Promotoren bestehen nicht nur im Vorkommen von TATA- oder GC-Boxen sondern auch darin, für welche Transkriptionsfaktoren sie Bindestellen besitzen und in welchem Abstand diese sich zueinander befinden. Die existierenden Bindestellen für ein bestimmtes Protein stimmen in ihrer Sequenz nicht vollständig überein, es finden sich aber konservierte Folgen von mindestens 4 Basen, die durch das Einfügen von „Wobbles“, d. h. Positionen, an denen sich jeweils unterschiedliche Basen befinden, noch verlängert werden können. Des weiteren liegen diese Bindestellen in bestimmten Abständen zueinander vor.

Die Verteilung der DNA im Interphase-Chromatin, das den größten Teil des nuklearen Volumens einnimmt, unterliegt jedoch einer ganz speziellen Ordnung. So ist die DNA an mehreren Stellen an die nukleare Matrix, eine filamentöse Struktur an der Innenseite der nuklearen Membran, angeheftet. Diese Regionen bezeichnet man als matrix attachment regions (MAR) oder scaffold attachment regions (SAR). Das Anheften hat wesentlichen Einfluß auf die Transkription bzw. die Replikation. Diese MAR-Fragmente weisen keine konservativen Sequenzen auf, bestehen allerdings zu 70% aus A bzw. T und liegen in der Nähe von cis-agierenden Regionen, die die Transkription allgemein regulieren, und Topoisomerase II-Erkennungsstellen.



Neben Promotoren und Enhancern existieren weitere regulatorische Elemente für verschiedene Gene, sogenannte Insulators. Diese Insulators können z.B. die Wirkung des Enhancers auf den Promotor inhibieren, wenn sie zwischen Enhancer und Promotor liegen, oder aber, zwischen Heterochromatin und einem Gen gelegen, das aktive Gen vor dem Einfluß des Heterochromatins schützen. Beispiele für solche Insulators sind: 1. sogenannte LCR (locus control regions), welche aus mehreren gegenüber DNAase I hypersensitiven Stellen besteht; 2. bestimmte Sequenzen wie SCS (specialized chromatin structures) bzw. SCS', 350 bzw. 200 bp lang und hoch-resistent gegen Degradierung durch DNAase I und auf beiden Seiten von hypersensitiven Stellen flankiert (Abstand je 100 bp). An scs' bindet das Protein BEAF-32. Diese Insulators können auf beiden Seiten des Gens liegen.

Eine Übersicht über den Stand der Technik in der Oligomer Array Herstellung läßt sich auch einer im Januar 1999 erschienen Sonderausgabe von Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999) und der dort zitierten Literatur entnehmen.

Patente, die sich allgemein auf die Verwendung von Oligomer Arrays und photolithographisches Maskendesign beziehen, sind z. B. US-A 5,837,832, US-A 5,856,174, WO-A 98/27430 und US-A 5,856,101. Zudem existieren einige Stoff- und Verfahrenspatente, welche die Verwendung photolabiler Schutzgruppen an Nukleosiden einschränken, so z. B. WO-A98/39348 und US-A 5,763,599.

Matrix-assistierte Laser Desorptions/Ionisations Massenspektrometrie (MALDI) ist eine neue, sehr leistungsfähige Entwicklung für die Analyse von Biomolekülen (Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10.000 daltons. Anal. Chem. 60: 2299-2301). Ein Analytmolekül wird in eine im UV absorbierende Matrix eingebettet. Durch einen kurzen Laserpuls wird die Matrix ins Vakuum verdampft und das Analyt so unfragmentiert in die Gasphase befördert. Eine angelegte Spannung beschleunigt die Ionen in ein feldfreies Flugrohr. Auf Grund ihrer verschiedenen Massen werden Ionen unterschiedlich stark beschleunigt. Kleinere Ionen erreichen den Detektor früher als größere und die Flugzeit wird in die Masse der

Ionen umgerechnet.

Für die Abtastung eines immobilisierten DNA-Arrays sind vielfach fluoreszent markierte Sonden verwendet worden. Besonders geeignet sind für die Fluoreszenzmarkierung ist das einfache Anbringen von Cy3 und Cy5 Farbstoffen am 5'OH der jeweiligen Sonde. Die Detektion der Fluoreszenz der hybridisierten Sonden erfolgt beispielsweise über ein Konfokalmikroskop. Die Farbstoffe Cy3 und Cy5 sind, neben vielen anderen, kommerziell erhältlich.

Um die erwartete Anzahl von amplifizierten Fragmenten ausgehend von einer beliebigen Templat-DNA und zweien nicht für jeweils eine bestimmte Position spezifischen Primern zu berechnen, muß ein statistisches Modell über den Aufbau des Genoms zu Grunde gelegt werden.

Wir geben hier die Berechnung für drei Modelle an, beziehen uns allerdings in diesem Patent auf die in Modell 3 beschriebene Methode.

Modell 1:

Im einfachsten Fall wird angenommen, daß ein primärer DNA-Strang eine Zufallsfolge von vier gleich häufig vorkommenden Basen ist. Damit ergibt sich als Wahrscheinlichkeit, daß sich für einen beliebiger Primer *PrimA* (der Länge *k*) an einer gegebenen Stelle im Genom eine perfekte Basenpaarung ergibt:

$$P_o(PrimA) = 0.25^k \quad (\text{Modell 1 für DNA})$$

(diese Wahrscheinlichkeit ist für den sense- und anti-sense-Strang der DNA gleich)

Bei einer Bisulfitbehandlung der DNA werden diejenigen Cytosine durch Uracil ersetzt, die nicht zu einem methylierten CG gehören. Das Basenpaarungsverhalten des Uracils entspricht dem des Thymins. Da CG in der DNA sehr selten sind (unter zwei Prozent), kann die statistische Häufigkeit der Cs nach der Bisulfitbehandlung vernachlässigt werden. Die Wahrscheinlichkeit, daß sich für einen Primer *PrimB* (Länge *k*, davon *a* As,

*t* Ts, *g* Gs und *c* Cs) auf bisulfitbehandelter DNA eine perfekte Basenpaarung ergibt, ist unterschiedlich für einen mit Bisulfit behandelten Strang und den zugehörigen anti-sense Strang:

$$P_{Is}(PrimB)=0.5^a * 0.25^t * 0.25^c * 0^g \quad (\text{Modell 1 für Bisulfit-DNA-Strang})$$

$$P_{Ia}(PrimB)=0.25^a * 0.5^t * 0^c * 0.25^g \quad (\text{Modell 1 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

(wenn der Primer C oder G enthält, wird somit einer der Wahrscheinlichkeitswerte 0).

#### Modell 2:

Zählungen der Basenhäufigkeiten der DNA ergeben, daß die vier Basen in der DNA nicht gleichverteilt sind. Entsprechend kann man aus DNA-Datenbanken folgende Häufigkeiten (Wahrscheinlichkeiten für ein Vorkommen) der Basen ermitteln.

$$P_{DNA}(A)=0.2811$$

$$P_{DNA}(T)=0.2784$$

$$P_{DNA}(C)=0.2206$$

$$P_{DNA}(G)=0.2199$$

Als Grundlage für diese Statistik (und die folgenden für Modell 2 und 3) dienen ca. 6% des Genoms vom Homo Sapiens aus High Throughput Sequencing Projekten (Datenbank "htgs" vom NIH/NCBI vom 6.9.1999). Die Gesamtmenge der Daten beträgt mehr als  $1.5 \times 10^8$  Basenpaare, was einem Schätzfehler für die Einzelwahrscheinlichkeiten kleiner  $10^{-5}$  entspricht.

Mit Hilfe dieser Werte läßt sich das Modell 1 verbessern.

Damit ist die Wahrscheinlichkeit, daß sich für einen Primer *PrimC* (Länge *k*, davon *a* As, *t* Ts, *g* Gs und *c* Cs) eine perfekte Basenpaarung ergibt:



$$P_2(PrimC) = P_{DNA}(T)^a * P_{DNA}(A)^t * P_{DNA}(C)^g * P_{DNA}(G)^c \quad (\text{Modell 3 für DNA})$$

Für den mit Bisulfit behandelten Strang ergeben sich folgende Wahrscheinlichkeiten unter der Annahme, daß alle CpG-Positionen methyliert sind (man erhält eine gleiche Statistik für die Bisulfitbehandlung des DNA-sense- und des DNA-antisense-Stranges):

$$P_{bDNA}(A) = 0.2811$$

$$P_{bDNA}(C) = 0.0140$$

$$P_{bDNA}(G) = 0.2199$$

$$P_{bDNA}(T) = 0.4850$$

Damit ergibt sich als Wahrscheinlichkeit, daß sich für einen Primer *PrimD* (Länge *k*, davon *a* As, *t* Ts, *g* Gs und *c* Cs) eine perfekte Basenpaarung ergibt:

$$P_{2s}(PrimD) = P_{bDNA}(T)^a * P_{bDNA}(A)^t * P_{bDNA}(C)^g * P_{bDNA}(G)^c \quad (\text{Modell 3 für Bisulfit-DNA-Strang})$$

$$P_{2a}(PrimD) = P_{bDNA}(A)^a * P_{bDNA}(T)^t * P_{bDNA}(G)^g * P_{bDNA}(C)^c \quad (\text{Modell 3 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

Modell 3:

Wesentliche Schätzfehler in Modell 2 ergeben sich vor allem bei der mit Bisulfit behandelten DNA aus der Tatsache, daß C nur noch im Kontext CG auftreten kann. Modell 3 berücksichtigt diese Eigenschaft und nimmt an, daß die primäre DNA eine Zufallsfolge mit Abhängigkeit direkt benachbarter Basen ist (Markov-Kette erster Ordnung). Die empirisch aus der Datenbank (vollständig methyliert; mit Bisulphit behandelt) ermittelten paarweisen Basenwahrscheinlichkeiten ergeben sich gleich für beide DNA-Stränge als  $P_{bDNA}(von,nach)$  aus der folgenden Tabelle:

<b>Von\nach</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

$$P_{bDNA}(A)=0.2811$$

$$P_{bDNA}(C)=0.0140$$

$$P_{bDNA}(G)=0.2199$$

$$P_{bDNA}(T)=0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge)  $P_{rbDNA}(von; nach)$

<b>Von\nach</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A)=0.4850 ;$$

$$P_{rbDNA}(C)=0.2199$$

$$P_{rbDNA}(G)=0.0140$$

$$P_{rbDNA}(T)=0.2811$$

Damit hängt die Wahrscheinlichkeit, daß sich für einen Primer *PrimE* (mit der Basenfolge  $B_1 B_2 B_3 B_4 \dots$ ; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(PrimE) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots \quad (\text{Modell 3 für})$$

Bisulfit-DNA-Strang)

$$P_{3a}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots \quad (\text{Modell 3 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

Berechnung der Anzahl der zu erwartenden amplifizierten Fragmente

Die mit Bisulfit behandelte DNA wird unter Benutzung einer Anzahl Primer amplifiziert. Aus Sicht des Modells besteht die DNA aus je einem sense- und einem anti-sense-Strang der Länge  $N$  Basen (alle Chromosomen werden hier zusammengefaßt). Für einen Primer  $Prim$  ist zu erwarten, daß er auf dem sense-Strang

$$N * P_s(Prim)$$

perfekte Basenpaarungen ergibt - für diese Berechnung können die Funktionen  $P_{1s}$ ,  $P_{2s}$  oder  $P_{3s}$  von Modell 1, 2 oder 3 eingesetzt werden, je nach gewünschter Abschätzungsgüte. Werden mehrere Primer ( $PrimU$ ,  $PrimV$ ,  $PrimW$ ,  $PrimX$ , etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) = & P_s(PrimU) \\ & + (1 - P_s(PrimU)) P_s(PrimV) \\ & + (1 - P_s(PrimU)) (1 - P_s(PrimV)) P_s(PrimW) \\ & + (1 - P_s(PrimU)) (1 - P_s(PrimV)) (1 - P_s(PrimW)) P_s(PrimX) \\ & + \dots \end{aligned}$$

Und damit als Anzahl der zu erwartenden perfekten Basenpaarungen mit irgendeinem der Primer

$$N * P_s(Primers)$$

Für die Bestimmung von  $P_a(\text{Primers})$  auf dem anti-sense-Strang werden die analogen Gleichungen verwendet. Ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge  $M$  ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet. Die Wahrscheinlichkeit dafür ist

$$P_a(\text{Primers}) \sum_{i=0}^{M-2} (1 - P_a(\text{Primers}))^i$$

Für große  $M$  und kleine  $P_a(\text{Primers})$  kann dieses durch folgenden Ausdruck berechnet werden:

$$\frac{1 - P_a(\text{Primers})}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1]$$

Für die Gesamtzahl  $F$  der Fragmente, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit:

$$F = N * P_s(\text{Primers}) \frac{(1 - P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] \\ + N * P_a(\text{Primers}) \frac{(1 - P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1]$$

Diese Methode liefert einen präzisen Erwartungswert für die Vorhersage der Anzahl der Bindungssites bestimmter Sequenzen an ein beliebiges zuvor mit Bisulfit behandeltes genomisches DNA Fragment. Sie dient hier als Grundlage für die Berechnung der statistisch erwarteten Anzahl von Amplifikaten in einer PCR-Reaktion ausgehend von zwei Primersequenzen und einer DNA der Länge  $N$ , wobei nur die Amplifikate berücksichtigt werden, die eine Anzahl von  $M$  Nukleotiden nicht überschreiten. In diesem Patent wird davon ausgegangen, daß  $M$  den Wert 2000 hat.

Die bekannten Verfahren für den Nachweis von Cytosin Methylierungen in genomischer DNA sind prinzipiell nicht so ausgelegt, daß eine Vielzahl von Zielregionen im zu untersuchenden Genom gleichzeitig erfaßt werden. Aufgabe der vorliegenden Erfindung ist es, ein Verfahren zu schaffen, mit dem eine Probe genomischer DNA gleichzeitig an mehreren Positionen gleichzeitig auf Cytosin Methylierung hin untersucht werden kann.

Die Aufgabe wird durch die kennzeichnenden Merkmale des Anspruchs 1 gelöst. Vorteilhafte Weiterbildungen der Merkmal sind in den abhängigen Ansprüchen gekennzeichnet.

Im Unterschied zu anderen Verfahren kann nach chemischer Vorbehandlung der DNA durch Verwendung entsprechend angepaßter Primerpaare eine Amplifikation von vielen Zielregionen gleichzeitig erfolgen. Dabei ist es nicht unbedingt notwendig den Sequenzkontext aller dieser Zielregionen vorab zu kennen, da in vielen Fällen, wie nachfolgend auch beispielhaft aufgeführt, Konsensussequenzen aus der Sequenzierung verwandter Zielregionen bekannt sind, die wie unten beschrieben für das Design für bestimmte Zielregionen spezifischer oder selektiver Primerpaare eingesetzt werden können. Das Verfahren ist dann erfolgreich angewandt, wenn die Amplifikation der chemisch vorbehandelten genomischen DNA mehr Fragmente bis maximal 2000 Basenpaare Länge als statistisch zu erwarten aus den jeweils zu untersuchenden Zielregionen liefert.

Dabei wird der statistische Erwartungswert für die Anzahl dieser Fragmente über die im Stand der Technik aufgeführten Formeln berechnet. Die Anzahl der im Amplifikationsschritt hergestellten Fragmente kann dagegen mittels einer beliebigen molekularbiologischen, chemischen oder physikalischen Methode nachgewiesen werden.

Für die Durchführung der erforderlichen statistischen Betrachtungen, die auch für die unten aufgeführten Ansprüche relevant sind, werden die folgenden Werte angenommen:

Das menschliche haploide Genom enthält 3 Milliarden Basenpaare und 100.000 Gene, die wiederum im Mittel eine 2000 Basenpaare lange mRNA codieren, die Gene inklusive der Introns sind durchschnittlich 15000 Basenpaare lang. Promotoren umfassen je Gen 1000 Basenpaare durchschnittlich. Ist daher der statistische Erwartungswert für die Anzahl der Amplifikate, die ausgehend von zwei Primern in transkribierten Sequenzen liegen, zu berechnen, so ist zunächst der Erwartungswert für das Gesamtgenom nach obiger Formel (Methode 3) zu berechnen und mit dem Anteil der transkribierten Sequenzen am Gesamtgenom zu berechnen. Analog wird für Teile eines beliebigen Genoms sowie für Promotoren und translatierte Sequenzen (mRNA codierend) vorgegangen.

Die vorliegende Erfindung beschreibt somit ein Verfahren zur parallelen Detektion des Methylierungszustandes genomischer DNA. Dabei sollen mehrere Cytosin-Methylierungen in einer DNA-Probe gleichzeitig analysiert werden. Dazu werden die folgenden Verfahrensschritte nacheinander ausgeführt:

Zuerst wird eine genomische DNA Probe derart chemisch behandelt, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden. Bevorzugt wird dazu die oben beschriebene Behandlung genomischer DNA mit Bisulfit (Hydrogensulfit, Disulfit) und anschließender alkalischer Hydrolyse verwendet, die zu einer Umwandlung nicht methylierter Cytosin-Nukleobasen in Uracil führt.

In einem zweiten Verfahrensschritt werden aus der vorbehandelten genomischen DNA mehr als zehn unterschiedliche Fragmente gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer amplifiziert, wobei mehr als doppelt so viele Fragmente als statistisch zu erwarten aus an der Genregulation beteiligten, transkribierten und/oder translatierten Sequenzen stammen. Dies kann mittels verschiedener Methoden erreicht werden.

In einer bevorzugten Variante des Verfahrens enthält mindestens eines der für die Amplifikation verwendeten Oligonukleotide weniger Nukleobasen als es statistisch für



eine sequenzspezifische Hybridisierung an die chemisch behandelte genomische DNA Probe erforderlich wäre, was zur Amplifikation mehrerer Fragmente gleichzeitig führen kann. Dabei ist die Gesamtzahl der in diesem Oligonukleotid enthaltenen Nukleobasen kleiner als 17. In einer besonders bevorzugten Variante des Verfahrens ist die Anzahl der in diesem Oligonukleotid enthaltenen Nukleobasen kleiner als 14.

In einer weiteren, bevorzugten Variante des Verfahrens werden für die Amplifikation mehr als 4 Oligonukleotide mit unterschiedlicher Sequenz gleichzeitig in einem Reaktionsgefäß verwendet. In einer besonders bevorzugten Varianten werden zur Herstellung eines komplexen Amplifikates mehr als 26 verschiedene Oligonukleotide gleichzeitig verwendet. In einer besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl, wie statistisch zu erwarten, aus an der Regulation von Genen beteiligten Genomabschnitten, z.B. Promotoren und Enhancern, stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre. In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, die in mindestens einer Zelle des jeweiligen Organismus in mRNA transkribiert werden, oder aber aus nach der Transkription in mRNA gesplizeten Genomabschnitten (Exons), als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, welche für Teile einer oder mehrerer Genfamilien kodieren, oder aber sie stammen aus Genomabschnitten, welche für sogenannte „matrix attachment sites“ (MARs)-charakteristische Sequenzen enthalten, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, welche als sogenannte „boundary elements“ die Verpackungsdichte des Chromatins organisieren, oder aber sie stammen aus multiple drug resistance gene“ (MDR)-

Promotoren oder kodierenden Regionen, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren, besonders bevorzugten Variante des Verfahrens werden zur Amplifikation der beschriebenen Fragmente zwei Oligonukleotide oder zwei Klassen von Oligonukleotiden verwendet, von denen eines oder eine Klasse außer im Kontext CpG oder CpNpG zwar die Base C enthalten kann, nicht aber die Base G und von denen das andere oder die andere Klasse außer im Kontext CpG oder CpNpG zwar die Base G, nicht aber die Base C enthalten kann.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide durchgeführt, von denen eines eine vier bis sechzehn Basen lange Sequenz enthält, die zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, an welches einer der Faktoren

AhR/Arnt	aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator
Arnt	aryl hydrocarbon receptor nuclear translocator
AML-1a	CBFA2; core-binding factor, runt domain, alpha subunit 2 (acute myeloid leukemia 1; aml1 oncogene)
AP-1	activator protein-1 (AP-1); Synonyme: c-Jun
C/EBP	CCAAT/enhancer binding protein
C/EBPalpha	CCAAT/enhancer binding protein (C/EBP), alpha
C/EBPbeta	CCAAT/enhancer binding protein (C/EBP), beta
CDP	CUTL1; cut (Drosophila)-like 1 (CCAAT displacement protein)
CDP	CUTL1; cut (Drosophila)-like 1 (CCAAT displacement protein)
CDP CR1	complement component (3b/4b) receptor 1
CDP CR3	complement component (3b/4b) receptor 3
CHOP-C/EBPalpha	DDIT; DNA-damage-inducible transcript 3/CCAAT/enhancer binding protein (C/EBP), alpha
c-Myc/Max	avian myelocytomatosis viral oncogene/MYC-ASSOCIATED FACTOR X
CREB	cAMP responsive element binding protein
CRE-BP1	CYCLIC AMP RESPONSE ELEMENT-BINDING PROTEIN 2, CREB2, CREBP1; now ATF2; activating transcription factor 2
CRE-BP1/c-Jun	activator protein-1 (AP-1); Synonyme: c-Jun

CREB	MP responsive element binding protein
E2F	E2F transcription factor (originally identified as a DNA-binding protein essential E1A-dependent activation of the adenovirus E2 promoter)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
Egr-1	early growth response 1
Egr-2	early growth response 2 (Krox-20 (Drosophila) homolog)
ELK-1	ELK1, member of ETS (environmental tobacco smoke) oncogene family
Freac-2	FKHL6; forkhead (Drosophila)-like 6; FORKHEAD-RELATED ACTIVATOR 2; FREAC2
Freac-3	FKHL7; forkhead (Drosophila)-like 7; FORKHEAD-RELATED ACTIVATOR 3; FREAC3
Freac-4	FKHL8; forkhead (Drosophila)-like 8; FORKHEAD-RELATED ACTIVATOR 4; FREAC4
Freac-7	FKHL11; forkhead (Drosophila)-like 9; FORKHEAD-RELATED ACTIVATOR 7; FREAC7
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-2	GATA-binding protein 2/Enhancer-Binding Protein GATA2
GATA-3	GATA-binding protein 3/Enhancer-Binding Protein GATA3
GATA-X	
HFH-3	FKHL10; forkhead (Drosophila)-like 10; FORKHEAD-RELATED ACTIVATOR 6; FREAC6
HNF-1	TCF1; transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
HNF-4	hepatocyte nuclear factor 4
IRF-1	interferon regulatory factor 1
ISRE	interferon-stimulated response element
Lmo2 complex	LIM domain only 2 (rhombotin-like 1)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
myogenin/NF-1	Myogenin (myogenic factor 4)/Neurofibromin 1; NEUROFIBROMATOSIS, TYPE I
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
NF-E2	NFE2; nuclear factor (erythroid-derived 2), 45kD
NF-kappaB (p50)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p50 subunit
NF-kappaB (p65)	nuclear factor of kappa light polypeptide gene enhancer in B-

	cells p65 subunit
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NRSF	NEURON RESTRICTIVE SILENCER FACTOR; REST; RE1-silencing transcription factor
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
P300	E1A (adenovirus E1A oncoprotein)-BINDING PROTEIN, 300-KD
P53	tumor protein p53 (Li-Fraumeni syndrome); TP53
Pax-1	paired box gene 1
Pax-3	paired box gene 3 (Waardenburg syndrome 1)
Pax-6	paired box gene 6 (aniridia, keratitis)
Pbx 1b	pre-B-cell leukemia transcription factor
Pbx-1	pre-B-cell leukemia transcription factor 1
RORalpha2	RAR-RELATED ORPHAN RECEPTOR ALPHA; RETINOIC ACID-BINDING RECEPTOR ALPHA
RREB-1	ras responsive element binding protein 1
SP1	simian-virus-40-protein-1
SP1	simian-virus-40-protein-1
SREBP-1	sterol regulatory element binding transcription factor 1
SRF	serum response factor (c-fos serum response element-binding transcription factor)
SRY	sex determining region Y
STAT3	signal transducer and activator of transcription 1, 91kD
Tal-1alpha/E47	T-cell acute lymphocytic leukemia 1/transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
TATA	cellular and viral TATA box elements
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
TCF11/MafG	v-maf musculoaponeurotic fibrosarcoma (avian) oncogene family, protein G
TCF11	Transcription Factor 11; TCF11; NFE2L1; nuclear factor (erythroid-derived 2)-like 1
USF	upstream stimulating factor
Whn	winged-helix nude

X-BP-1  
YY1

X-box binding protein 1 oder  
ubiquitously distributed transcription factor belonging to  
the GLI-Kruppel class of zinc finger proteins

bindet, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide oder zweier Klassen von Oligonukleotiden durchgeführt, von denen eines oder die eine Klasse die vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide oder zweier Klassen von Oligonukleotiden durchgeführt, von denen eines oder die eine Klasse eine der Sequenzen

TCGCGTGTA, TACACGCGA, TGTACGCGA, TCGCGTACA,  
TTGCGTGTT, AACACGCAA, GGTACGTAA, TTACGTACC,  
TCGCGTGTT, AACACGCGA, GGTACGCGA, TCGCGTACC,  
TTGCGTGTA, TACACGCAA, TGTACGTAA, TTACGTACA,  
TACGTG, CACGTA, TACGTG, CACGTA,

ATTGCGTGT, ACACGCAAT, GTACGTAAT, ATTACGTAC,  
ATTGCGTGA, TCACGCAAT, TTACGTAAT, ATTACGTAA,  
ATCGCGTGA, TCACGCGAT, TTACGCGAT, ATCGCGTAA,  
ATCGCGTGT, ACACGCGAT, GTACGCGAT, ATCGCGTAC,  
TGTGGT, ACCACA, ATTATA, TATAAT,

TGAGTTAG, CTAACCTCA, TTGATTTA, TAAATCAA,  
TGATTTAG, CTAAATCA, TTGAGTTA, TAACTCAA,

TTTGGT, ACCAAA, ATTAAT, TTAAAT,  
TGTGGA, TCCACA, TTTATA, TATAAA,  
TTTGGG, TCCAAA, TTATAA, TTATAA,  
TGTGGT, ACCACA, ATTATA, TATAAT,

ATTAT, ATAAT, GTAAT, ATTAC,  
ATTGT, ACAAT, GTAAT, ATTAC,

GAAAG, CTTTC, TTTTT, AAAAA,  
GTAAT, ATTAC, ATTGT, ACAAT,  
GAAAT, ATTTT, ATTTT, AAAAT,  
GTAAG, CTTAC, TTTGT, AAAAA,  
TTAATAATCGAT, ATCGATTATTAA, ATCGATTATTGG, CCAATAATCGAT,  
ATCGATTA, TAATCGAT, TAATCGAT, ATCGATTA,

ATCGATCGG, CCGATCGAT, TCGATCGAT, ATCGATCGA,  
ATCGATCGT, ACGATCGAT, GCGATCGAT, ATCGATCGC,

TATCGATA, TATCGATA, TATCGGTG, CACCGATA,  
TATTAATA, TATTAATA, TATTGGTG, CACCAATA,

GTGTAATATTT, AAATATTACAC, GGGTATTGTAT, ATACAATACCC,  
GTGTAATTTTT, AAAAATTACAC, GGGGATTGTAT, ATACAATCCCC,  
ATGTAATTTTT, AAAAATTACAT, GGGGATTGTAT, ATACAATCCCC,  
ATGTAATATTT, AAATATTACAT, GGGTATTGTAT, ATACAATACCC,  
ATTACGTGGT, ACCACGTAAT, ATTACGTGGT, ACCACGTAAT,  
TGACGTAA, TTACGTCA, TTACGTTA, TAACGTAA,  
TGACGTTA, TAACGTCA, TGACGTTA, TAACGTCA,  
TTACGTAA, TTACGTAA, TTACGTAA, TTACGTAA,  
TGACGTTA, TAACGTCA, TAACGTTA, TAACGTTA,

TGACGT, ACGTCA, GCGTTA, TAACGC,  
TGACGT, ACGTCA, ACGTTA, TAACGT,  
TTTCGCGT, ACGCGAAA, GCGCGAAA, TTTCGCGC,  
TTTGCGGT, ACGCCAAA, GCGTTAAA, TTAAACGC,

TAGGTGTTA, TAACACCTA, TAATATTTG, CAAATATTA,  
TAGGTGTTT, AAACACCTA, GAATATTTG, CAAATATTC,

GTAGGTGG, CCACCTAC, TTATTTGT, ACAAATAA,  
GTAGGTGT, ACACCTAC, ATATTTGT, ACAAATAT,

TGCGTGGGCGG, CCGCCCACGCA, TCGTTTACGTA, TACGTAAACGA,  
TGCGTGGGCGT, ACGCCCACGCA, ACGTTTACGTA, TACGTAAACGT,

TGCGTAGGCGT, ACGCCTACGCA, ACGTTTACGTA, TACGTAAACGT,  
TGCGTAGGCGG, CCGCCTACGCA, TCGTTTACGTA, TACGTAAACGA,  
ATAGGAAGT, ACTTCCTAT, ATTTTTTGT, AAAAAAAT,



TCGGAAGT, ACTTCCGA, ATTTTCGG, CCGAAAAT,  
TCGGAAGT, ACTTCCGA, GTTTTCGG, CCGAAAAC,  
TCGGAAT, ATTTCCGA, ATTTTCGG, CCGAAAAT,  
TCGGAAT, ATTTCCGA, GTTTTCGG, CCGAAAAC,  
GTAAATAA, TTATTTAC, TTGTTTAT, ATAAACAA,  
GTAAATAAATA, TATTTATTTAC, TGTTTATTTAT, ATAAATAAACA,

AAAGTAAATA, TATTTACTTT, TGTTTATTTT, AAAATAAACA,  
AATGTAAATA, TATTTACATT, TGTTTATATT, AATATAAACA,  
TAAGTAAATA, TATTTACTTA, TGTTTATTTA, TAAATAAACA,  
TATGTAAATA, TATTTACATA, TGTTTATATA, TATATAAACA,

ATAAATA, TATTTAT, TGTTTAT, ATAAACA,  
ATAAATA, TATTTAT, TATTTAT, ATAAATA,  
GATA, TATC, TATT, AATA,

TAGATAA, TTATCTA, TTATTTG, CAAATAA,  
TTGATAA, TTATCAA, TTATTAG, CTAATAA,  
GATAA, TTATC, TTATT, AATAA,

GATG, CATC, TATT, AATA,

GATAG, CTATC, TTATT, AATAA,  
GATAAG, CTTATC, TTTATT, AATAAA,

TGTTTATTTA, TAAATAAACA, TAAATAAATA, TATTTATTTA,  
TGTTTGTTTA, TAAACAAACA, TAAATAAATA, TATTTATTTA,  
TATTTATTTA, TAAATAAATA, TAAATAAATA, TATTTATTTA,  
TATTTGTTTA, TAAACAAATA, TAAATAAATA, TATTTATTTA,

GTTAATGATT, AATCATTAAC, AATTATTAAT, ATTAATAATT,  
GTTAATTATT, AATAATTAAC, AATAATTAAT, ATTAATTATT,  
GTTAATTAAT, ATTAATTAAC, ATTAATTAAT, ATTAATTAAT,  
GTTAATGAAT, ATTCATTAAC, ATTTATTAAT, ATTAATAAAT,

TAAAGTTTA, TAAACTTTA, TGAATTTTG, CAAAATTCA,  
TAAAGGTTA, TAACCTTTA, TGATTTTGT, CAAAATCA,

AAAGTGAAATT, AATTTCACTTT, GGTTTTATTTT, AAAATAAAACC,  
AAAGCGAAATT, AATTTGCTTT, GGTTCGTTTT, AAAACGAAACC,

TAGTTTTATTTTTT, AAAAAATAAACTA, GGGAAAGTGAAATTG,  
CAATTTCACTTTCCC,  
TAGTTTTATTTTTT, AAAAAATAAACTA, GGAAAAGTGAAATTG,  
CAATTTCACTTTTCC,  
TAGTTTTTTTTTTT, AAAAAAAAAAACTA, GGAAAAGAGAAATTG,  
CAATTTCTCTTTTCC,

TAGTTTTTTTTTTTTT, AAAAAAAAAAAAACTA, GGGAAAGAGAAATTG,  
CAATTTCTCTTTCCC,  
TAGGTG, CACCTA, TATTTG, CAAATA,

TTTTAAAAATAATTTT, AAAATTATTTTTTAAAA, AGGGTTATTTTTTAGAG,  
CTCTAAAAATAACCCT,  
TTTTAAAAATAATTTT, AAAATTATTTTTTAAAA, GGAGTTATTTTTTAGAG,  
CTCTAAAAATAACTCC,  
TTTTAAAAATAATTTT, AAAATTATTTTTTAAAA, AGAGTTATTTTTTAGAG,  
CTCTAAAAATAACTCT,  
TTTTAAAAATAATTTT, AAAATTATTTTTTAAAA, GGGGTTATTTTTTAGAG,  
CTCTAAAAATAACCCC,

TGTTATTAAAAATAGAAA, TTTCTATTTTTTAATAACA,  
TTTTTATTTTTTAGTAATA, TATTACTAAAAATAAAAA,  
TGTTATTAAAAATAGAAT, ATTCTATTTTTTAATAACA,  
GTTTTATTTTTTAGTAATA, TATTACTAAAAATAAAAC,  
TTTGGTAT, ATACCAA, GTGTTAAA, TTAAACAC  
GGGA, TCCCC, TTTTT, AAAAA,

TAGGGG, CCCCTA, TTTTTA, TAAAAA,  
GAGGGG, CCCCTC, TTTTTT, AAAAAA,  
TGTTGAGTTAT, ATAACCAACA, ATGATTTAGTA, TACTAAATCAT,  
TGTTGATTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAACTCAC,  
TGTTGAGTTAT, ATAACCAACA, ATGATTTAGTA, TACTAAATCAT,  
TGTTGATTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAACTCAC,  
GGGGATTTTT, AAAAATCCCC, GGGAATTTTT, AAAAATTCCC,  
GGGGATTTTT, AAAAATCCCC, GGGGATTTTT, AAAAATCCCC,  
GGGGATTTTT, AAAAATCCCC, GGAAATTTTT, AAAAATTCC,  
GGGAATTTTT, AAAAATTCCC, GGAAATTTTT, AAAAATTCC,  
GGGAATTTTT, AAAAATTCCC, GGAAATTTTT, AAAAATTCC,  
GGGATTTTTT, AAAAAATCCC, GGAAAGTTTT, AAAACTTTCC,  
GGGAATTTTT, AAAAATTCCC, GGGAATTTTT, AAAAATTCCC,  
GGGATTTTTT, AAAAAATCCC, GGGAAGTTTT, AAAACTTTCC,  
GGGATTTTTTA, TAAAAAATCCC, TGGAAAGTTTT, AAAACTTTCCA,  
TTTAGTATTACGGATAGAGGT, ACCTCTATCCGTAATACTAAA,  
GTTTTGTTTCGTGGTGTGAA, TTCAACACCACGAACAAAAAC,  
TTTAGTATTACGGATAGAGTT, AACTCTATCCGTAATACTAAA,  
GGTTTTGTTTCGTGGTGTGAA, TTCAACACCACGAACAAAACC,  
TTTAGTATTACGGATAGCGTT, AACGCTATCCGTAATACTAAA,  
GGCGTTGTTTCGTGGTGTGAA, TTCAACACCACGAACAACGCC,  
TTTAGTATTACGGATAGCGGT, ACCGCTATCCGTAATACTAAA,  
GTCGTTGTTTCGTGGTGTGAA, TTCAACACCACGAACAACGAC,

ATATGTAAAT, ATTTACATAT, ATTTGTATAT, ATATACAAAT,  
TTATGTAAAT, ATTTACATAA, ATTTGTATAA, TTATACAAAT,

GAATATTTA, TAAATATTC, TGAATATTT, AAATATTCA,

GAATATGTA, TACATATTC, TGTATATTT, AAATATACA,  
ATAAT, ATTAT, ATTAT, ATAAT,  
GTAAT, ATTAC, ATTAT, ATAAT,  
AATGTAAAT, ATTTACATT, ATTTGTATT, AATACAAAT,  
ATTTGTATATT, AATATACAAAT, GGTATGTAAAT, ATTTACATACC,  
ATTTGTATATT, AATATACAAAT, AATATGTAAAT, ATTTACATATT,  
ATTTGTATATT, AATATACAAAT, AGTATGTAAAT, ATTTACATACT,  
ATTTGTATATT, AATATACAAAT, GATATGTAAAT, ATTTACATATC,  
AGGAGT, ACTCCT, ATTTTT, AAAAAT,  
GGGAGT, ACTCCC, ATTTTT, AAAAAT,  
GGATATGTTCTGGGTATGTTT, AAACATACCCGAACATATCC,  
GGATATGTTCTGGGTATGTTT, AAACATACCCGAACATATCC,  
GGATATGTTCTGGGTATGTTT, AAACATACCCGAACATATCC,  
AGATATGTTCTGGGTATGTTT, AAACATACCCGAACATATCT,  
TCGTTTCGTTTTAGATAT, ATATCTAAAACGAAACGA,  
ATATTTAGAGCGGAACGG, CCGTTCCGCTCTAAATAT,  
CGTTACGGTT, AACCGTAACG, AATCGTGACG, CGTCACGATT,  
CGTTACGGTT, AACCGTAACG, GATCGTGACG, CGTCACGATC,  
CGTTACGTTT, AACCGTAACG, AAGCGTGACG, CGTCACGCTT,  
CGTTACGTTT, AACCGTAACG, GAGCGTGACG, CGTCACGCTC,  
TTTACGTATGA, TCATACGTAAA, TTATGCGTGAA, TTCACGCATAA,  
TTTACGTTTGA, TCAAACGTAAA, TTAAGCGTGAA, TTCACGCTTAA,  
TTTACGTTTTA, TAAACGTAAA, TGAAGCGTGAA, TTCACGCTTCA,  
TTTACGTATTA, TAATACGTAAA, TGATGCGTGAA, TTCACGCATCA,  
AATTAATTAA, TTAATTAATT, TTGATTGATT, AATCAATCAA,  
TATTAATTAA, TTAATTAATA, TTGATTGATG, CATCAATCAA,  
TAATTAT, ATAATTA, ATGATTG, CAATCAT,  
TAGGTTA, TAACCTA, TGATTTA, TAAATCA,  
TTTTAAATATTTTT, AAAAATATTTAAAA, GGGGGTGTTTGGGG,  
CCCCAAACACCCCC,  
TTTTAAATTATTTTT, AAAATAATTTAAAA, GGGGTGGTTTGGGG,  
CCCCAAACCACCCC,  
TTTTAAATTTTTTT, AAAAAAATTTAAAA, GGGGGGGTTTGGGG,  
CCCCAAACCCCCC,  
TTTTAAATAATTTTT, AAAATTATTTAAAA, GGGGTTGTTTGGGG,  
CCCCAAACAACCCC,  
GAGGCGGGG, CCCC GCCTC, TTTCGTTTT, AAAACGAAA,

GAGGTAGGG, CCCTACCTC, TTTTGTTTT, AAAACAAAA,  
AAGGCGGGG, CCCC GCCTT, TTTCGTTTT, AAAACGAAA,  
AAGGTAGGG, CCCTACCTT, TTTTGTTTT, AAAACAAAA,

GGGGGCGGGGT, ACCCCGCCCCC, ATTTTCGTTTT, AAAACGAAAT,  
GGGGGCGGGGT, ACCCCGCCCCC, GTTTTCGTTTT, AAAACGAAAC,  
TATTATTTTAT, ATAAAATAATA, GTGGGGTGATA, TATCACCCAC,  
GATTATTTTAT, ATAAAATAATC, GTGGGGTGATT, AATCACCCAC,

ATTACGTGAT, ATCACGTAAT, ATTACGTGAT, ATCACGTAAT,  
ATTACGTGAT, ATCACGTAAT, GTTACGTGAT, ATCACGTAAC,

TTTTATATGG, CCATATAAAA, TTATATAAGG, CCTTATATAA,  
TTATATATGG, CCATATATAA, TTATATATGG, CCATATATAA,  
AAATAAT, ATTATTT, GTTGTTT, AAACAAC,  
AAATTAA, TTAATTT, TTAGTTT, AAATAA,  
AAATTAT, ATAATTT, GTAGTTT, AAATAA,  
AAATAAA, TTTATTT, TTTGTTT, AAATAA,

ATTTTTCGGAAATG, CATTTCCGAAAAAT, TATTTTCGGGAAAT,  
ATTTCCCGAAAATA,  
ATTTTTCGGAAATG, CATTTCCGAAAAAT, TATTTTCGGGAAAT,  
ATTTCCCGAAAATA,  
ATTTTCGGGAAATG, CATTTCCCGAAAAT, TATTTTTCGGAAAT,  
ATTTCCGAAAAATA,  
ATTTTCGGGAAGTG, CACTTCCCGAAAAT, TATTTTTCGGAAAT,  
ATTTCCGAAAAATA,

AATAGATGTT, AACATCTATT, AATATTTGTT, AACAAATATT,  
AATAGATGGT, ACCATCTATT, ATTATTTGTT, AACAAATAAT,

GTATAAATA, TATTTATAC, TATTTATAT, ATATAAATA,  
GTATAAATG, CATTTATAC, TATTTATAT, ATATAAATA,  
GTATAAAAA, TTTTATAC, TTTTATAT, ATATAAAAA,  
GTATAAAAG, CTTTATAC, TTTTATAT, ATATAAAAA,  
TTATAAATA, TATTTATAA, TATTTATAG, CTATAAATA,  
TTATAAATG, CATTTATAA, TATTTATAG, CTATAAATA,  
TTATAAAAA, TTTTATAA, TTTTATAG, CTATAAAAA,  
TTATAAAAG, CTTTATAA, TTTTATAG, CTATAAAAA,  
GGGGGTGACGTA, TACGTCAACCCCC, TCGTTAATTTTT,  
AAAAATTAACGCA,  
GGGGGTGACGTA, TACGTCAACCCCC, TACGTTAATTTTT,  
AAAAATTAACGTA,

TGACGTATATTTTT, AAAAATATACGTCA, GGGGATATGCGTTA,  
TAACGCATATCCCC,  
TGACGTATATTTTT, AAAAATATACGTCA, GGGGGTATGCGTTA,  
TAACGCATACCCCC,

ATGATTTAGTA, TACTAAATCAT, TGTTGAGTTAT, ATAACTCAACA,  
GTTAT, ATAAC, ATGAT, ATCAT,

TTACGTGA, TCACGTAA, TTACGTGG, CCACGTAA,  
TTACGTGG, CCACGTAA, TTACGTGG, CCACGTAA,  
TTACGTGG, CCACGTAA, TTACGTGA, TCACGTAA,  
TTACGTGA, TCACGTAA, TTACGTGA, TCACGTAA,  
GACGTT, AACGTC, AGCGTT, AACGCT,

TGACGTGT, ACACGTCA, ATACGTTA, TAACGTAT,  
TGACGTGG, CCACGTCA, TTACGTTA, TAACGTAA,  
CGGTTATTTTG, CAAAATAACCG, TAAGATGGTCG oder CGACCATCTTA

enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer besonders bevorzugten Variante des Verfahrens enthalten die zur Amplifikation verwendeten Oligonukleotide außer den oben definierten Konsensussequenzen mehrere Positionen, an denen entweder irgendeine der drei Basen G, A und T oder irgendeine der Basen C, A und T vorhanden sein kann.

In einer besonders bevorzugten Variante des Verfahrens enthalten die zur Amplifikation verwendeten Oligonukleotide außer einer der oben beschriebenen Konsensussequenzen nur maximal zusätzlich so viele weitere Basen, wie es zur gleichzeitigen Amplifikation von mehr als einhundert verschiedenen Fragmenten pro Reaktion aus der chemisch wie oben behandelten DNA erforderlich ist.

In einem dritten Verfahrensschritt wird nun der Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide untersucht.

In einer besonders bevorzugten Variante des Verfahrens erfolgt die Analyse durch Hybridisierung der bereits in der Amplifikation mit einem Fluoreszenzmarker versehenen Fragmente an einen Oligonukleotid- Array (DNA Chip). Der Fluoreszenzmarker kann entweder über die verwendeten Primer oder aber durch ein fluoreszenzmarkiertes Nukleotid (z. B. Cy5-dCTP, kommerziell von Amersham-Pharmacia erhältlich) eingeführt werden.

Dabei hybridisieren komplementäre Fragmente an die jeweiligen auf der Chipoberfläche immobilisierten Oligomere, nicht komplementäre Fragmente werden in einem oder mehreren Waschschritten entfernt. Die Fluoreszenz an den jeweiligen Hybridisierungsorten auf dem Chip erlaubt dann den Rückschluß auf den Sequenzkontext der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.

In einer weiteren bevorzugten Variante des Verfahrens werden die amplifizierten Fragmente auf einer Oberfläche immobilisiert und anschließend eine Hybridisierung mit einer kombinatorischen Bibliothek von unterscheidbaren Oligonukleotid- oder PNA-Oligomer-Sonden durchgeführt. Wiederum werden nicht komplementäre Sonden durch einen oder mehrere Waschschriffe entfernt. Die hybridisierten Sonden werden entweder über ihre Fluoreszenzmarker detektiert oder in einer weiteren besonders bevorzugten Variante des Verfahrens mittels Matrix-assistierter Laser-Desorptions/Ionisations Massenspektrometrie (MALDI-MS) anhand ihrer eindeutigen Masse nachgewiesen. Dabei werden die Sondenbibliotheken derart synthetisiert, daß die Masse eines jeden Bestandteils eindeutig seiner Sequenz zugeordnet werden kann.

Die Amplifikate können zudem in einer weiteren bevorzugten Variante des Verfahrens hinsichtlich ihrer durchschnittlichen Größe durch Veränderung der Kettenverlängerungszeiten im Amplifikationsschritt beeinflußt werden. Da hier vorwiegend kleinere Fragmente (ca. 200-500 Basenpaare) untersucht werden, ist eine Verkürzung der Kettenverlängerungsschritte z. B. einer PCR sinnvoll.

In einer weiteren bevorzugten Variante des Verfahrens werden die Amplifikate durch



Gelelektrophorese aufgetrennt, und die Fragmente im gewünschten Größenbereich werden vor Ihrer Analyse ausgeschnitten. In einer weiteren besonders bevorzugten Variante werden die aus dem Gel ausgeschnittenen Amplifikate unter Verwendung des gleichen Satzes an Primern erneut amplifiziert. Dabei können dann nur noch Fragmente der gewünschten Größe entstehen, da Andere als Templat nicht mehr verfügbar sind.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Kit, enthaltend mindestens zwei Primerpaare, Reagenzien und Hilfsstoffe für die Amplifikation und/oder Reagenzien und Hilfsmittel für die chemische Behandlung und/oder eine kombinatorische Sondenbibliothek und/oder einen Oligonukleotid-Array (DNA-Chip), soweit sie für die Durchführung des erfindungsgemäßen Verfahrens erforderlich oder dienlich sind.

Die folgenden Beispiele erläutern die Erfindung.

Beispiele:

Beispiel 1:

Primer zur bevorzugten Amplifikation von CG reichen Regionen im Humangenom

Bei den CG reichen Regionen im Humangenom handelt es sich um sogenannte CpG-islands, die eine regulatorischen Funktion besitzen. Wir definieren CpG Islands derart, dass sie mindestens 500 bp umfassen sowie einen GC-Gehalt von >50% aufweisen, ausserdem ist der Quotient  $CG/GC > 0,6$ . Unter diesen Bedingungen liegen 16 Mb als CpG Islands vor. Damit liegen etwa 0,5 % der Genomsequenz in diesen CpG islands, wenn man auch noch jeweils eine Region bis 1000 bp downstream zusätzlich betrachtet. Dieser Überlegung liegen Daten aus der Ensembl Database vom 31.10.00, Quelle Sanger Centre, zugrunde. Die dort verfügbare Sequenz umfasste ca. 3,5 GB, und für die Berechnungen wurden die Repeats maskiert.

Statistisch wäre es bei 12meren zu erwarten, dass sie nur 0,005 mal so häufig an eine der CG-reichen Regionen hybridisieren wie an eine andere beliebige Region im

Genom. Es wurden nun Primer gefunden, welche 1,8 mal häufiger an eine CG reiche Region binden. Zudem ergibt sich mit den entsprechend gefundene Reverse Primer nahezu eine Spezifität für diese CpG islands.

In diesem Beispiel sind die Primer AGTAGTAGTAGT (Seq. ID 1) AAAACAAAAACC (Seq. ID 2) und alternativ AGTAGTAGTAGT (Seq. ID 19) und ACAAAAACATAA (seq. ID 20). Das erste Primerpaar führt mindestens zu den Amplifikaten Seq. ID 3 bis 18, das zweite Primerpaar zu den Amplifikaten der Seq. ID 21 bis 31.

#### Beispiel 2:

Berechnung der Vorhersage der Anzahl von Amplifikaten in Genomischen Regionen.

Gemäß Anspruch 8 im Patent wird gezeigt mehr als doppelt so viele Amplifikate erstellen zu können, als es statistisch zu erwarten wäre nach Formel 1.

$$F = N * P_s(Primers) \frac{(P_a(Primers))}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] + N * P_a(Primers) \frac{(P_s(Primers))}{\log(1 - P_s(Primers))} [(1 - P_s(Primers))^M - 1] \quad \text{Formel 1.}$$

F gibt dabei die Anzahl der Vorhergesagten Amplifikate an, die zu erwarten sind, wenn man N Basen als Datenbasis aus dem Genom betrachtet. P ist die jeweilige Wahrscheinlichkeit für die Hybridisierung eines Primeroligonukleotids, getrennt nach Hybridisierung im Sense- und Antisense-Strang. M ist die maximal zulässige Länge der zu erwartenden Amplifikate.

Die Wahrscheinlichkeit P wird bestimmt durch eine Markov Kette erster Ordnung. Dabei wird die Annahme gemacht, dass die DNA eine Zufallsfolge in Abhängigkeit benachbarter Basen ist. Für die Berechnung einer Markovkette sind die Übergangswahrscheinlichkeiten von benachbarten Basen notwendig. Diese wurden empirisch aus 12% des assemblierten humanen Genoms, das vollständig mit Bisulfit behandelt wurde, ermittelt und in Tabelle 1 zusammengefasst. In Tabelle 2 sind die Übergangswahrscheinlichkeiten für den entsprechenden komplementär reversen

Strang angegeben. Diese ergeben sich durch einfaches Vertauschen der Einträge aus der Tabelle 1.

Tabelle 1

<b>Von\nach</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

mit

$$P_{bDNA}(A)=0.2811$$

$$P_{bDNA}(C)=0.0140$$

$$P_{bDNA}(G)=0.2199$$

$$P_{bDNA}(T)=0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge)  $P_{rbDNA}(von; nach)$

Tabelle 2

<b>Von\nach</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A)=0.4850$$

$$P_{rbDNA}(C)=0.2199$$

$$P_{rbDNA}(G)=0.0140$$

$$P_{rbDNA}(T)=0.2811$$

Damit hängt die Wahrscheinlichkeit, dass sich für einen Primer *PrimE* (mit der

Basenfolge  $B_1 B_2 B_3 B_4 \dots$ ; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(PrimE) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots$$

(Bisulfit-DNA-Strang)

$$P_{3a}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots$$

(anti-sense-Strang zu einem Bisulfit-DNA-Strang);

für einen Primer  $Prim$  auf dem sense-Strang ergeben sich

$$N * P_s(Prim)$$

perfekte Basenpaarungen - Werden mehrere Primer ( $PrimU$ ,  $PrimV$ ,  $PrimW$ ,  $PrimX$ , etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) = & P_s(PrimU) \\ & + (1 - P_s(PrimU)) P_s(PrimV) \\ & + (1 - P_s(PrimU)) (1 - P_s(PrimV)) P_s(PrimW) \\ & + (1 - P_s(PrimU)) (1 - P_s(PrimV)) (1 - P_s(PrimW)) P_s(PrimX) \\ & + \dots \end{aligned}$$

( $PrimU$ ,  $PrimV$ ,  $PrimW$ ... sind hier verschiedene Primer mit unterschiedlichen Basenpaarungen)

und damit als Anzahl der zu erwartenden perfekten Basenpaarungen mit irgendeinem der Primer

$$N * P_s(Primers)$$

Für die Bestimmung von  $P_a(Primers)$  auf dem anti-sense-Strang werden die analogen Gleichungen verwendet.

Für das Beispiel mit zwei Primern (einem sense-Primer und einem antisense-Primer) ergeben sich folgende Wahrscheinlichkeiten:

$$P(\text{AGTAGTAGTAGT}) = 0.000000860027$$

$$P(\text{AACAAAACTAA}) = 0.000030005828$$

Auf den CpG-Islands, die insgesamt ca. 30.000.000 Basen enthalten, erwartet man eine Häufigkeit von Hybridisierungen für:

AGTAGTAGTAGT: 25.80 auf dem sense Strang

AACAAAAACTAA: 900.17 auf dem komplementär reversen Strang.

Auf den jeweils anderen Strängen können die Primer nicht hybridisieren, da auf dem sense-Strang durch die Bisulfitbehandlung keine Cs außerhalb des Kontextes CG auftreten und entsprechend komplementär auf dem antisense-Strang.

Ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge  $M$  ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet, die Wahrscheinlichkeit dafür ist

$$P_a(Primers) \sum_{i=0}^{M-2} (1 - P_a(Primers))^i ;$$

für große  $M$  und kleine  $P_a(Primers)$  wird dieses durch folgenden Ausdruck berechnet:

$$\frac{P_a(Primers)}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] ;$$

für die Gesamtzahl  $F$  der Amplifikate, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit

$$F = N * P_s(Primers) \frac{(P_a(Primers))}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] + N * P_a(Primers) \frac{(P_s(Primers))}{\log(1 - P_s(Primers))} [(1 - P_s(Primers))^M - 1] \quad \text{Formel 1}$$

Für das oben angegebene Beispiel ergeben sich für die CpG-Islands mit 30 Mega Basen 3.0498 Amplifikate. Wir können jedoch zeigen (siehe Beispiel 1), dass man mit Primern, die für bestimmte Regionen spezifisch sind, mehr als statistisch vorhergesagte Amplifikate erzeugen kann.

### Patentansprüche

1. Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA, dadurch gekennzeichnet, daß man folgende Schritte ausführt:
  - a) in einer genomischen DNA Probe wandelt man durch chemische Behandlung an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base um;
  - b) aus dieser chemisch behandelten genomischen DNA amplifiziert man mehr als zehn unterschiedliche Fragmente, die jeweils weniger als 2000 Basenpaare lang sind, gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer, wobei diese Primer jeweils Sequenzen aus an der Genregulation beteiligten und/oder transkribierten und/oder translatierten genomischen Sequenzen enthalten, wie sie nach einer Behandlung gemäß Schritt a) vorliegen würden;
  - c) man bestimmt den Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß man die chemische Behandlung mittels einer Lösung eines Bisulfits, Hydrogensulfits oder Disulfits durchführt.
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) verwendeten Oligonukleotide weniger Nukleobasen enthält als es statistisch für eine sequenzspezifische Hybridisierung an die chemisch behandelte genomische DNA Probe erforderlich wäre.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) des Anspruchs 1 verwendeten Oligonukleotide



kürzer als 18 Nukleobasen ist.

5. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) des Anspruchs 1 verwendeten Oligonukleotide kürzer als 15 Nukleobasen ist.
6. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß man in Schritt b) des Anspruchs 1 mehr als 4 verschiedene Oligonukleotide gleichzeitig für die Amplifikation verwendet.
7. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß man in Schritt b) des Anspruchs 1 mehr als 26 verschiedene Oligonukleotide gleichzeitig für die Amplifikation verwendet.
8. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als nach berechnet nach Formel 1 aus an der Regulation von Genen beteiligten Genomabschnitten, wie Promotoren und Enhancern, stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist wie berechnet nach Formel 1,

$$F = N * P_s(Primers) \frac{(P_a(Primers))}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] + N * P_a(Primers) \frac{(P_s(Primers))}{\log(1 - P_s(Primers))} [(1 - P_s(Primers))^M - 1]$$

Formel 1

wobei man die Berechnung wie folgt durchführt:

bei der mit Bisulfit behandelten DNA kann C nur noch im Kontext CG auftreten, so wird angenommen, daß die primäre DNA eine Zufallsfolge mit Abhängigkeit direkt benachbarter Basen ist (Markov-Kette erster Ordnung); die empirisch aus der Datenbank (vollständig methyliert; mit Bisulfit behandelt) ermittelten

paarweisen Basenwahrscheinlichkeiten ergeben sich gleich für beide DNA-Stränge als  $P_{bDNA}(von;nach)$  aus der folgenden Tabelle:

<b>Von\nach</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

mit

$$P_{bDNA}(A)=0.2811$$

$$P_{bDNA}(C)=0.0140$$

$$P_{bDNA}(G)=0.2199$$

$$P_{bDNA}(T)=0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge)  $P_{rbDNA}(von;nach)$

<b>Von\nach</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A)=0.4850$$

$$P_{rbDNA}(C)=0.2199$$

$$P_{rbDNA}(G)=0.0140$$

$$P_{rbDNA}(T)=0.2811$$

;damit hängt die Wahrscheinlichkeit, daß sich für einen Primer *PrimE* (mit der Basenfolge  $B_1 B_2 B_3 B_4 \dots$ ; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(PrimE) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots$$

(Bisulfit-DNA-Strang)

$$P_{3a}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots$$

(anti-sense-Strang zu einem Bisulfit-DNA-Strang);

für einen Primer *Prim* auf dem sense-Strang ergeben sich

$$N * P_s(Prim) ;$$

perfekte Basenpaarungen - Werden mehrere Primer (*PrimU*, *PrimV*, *PrimW*, *PrimX*, etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) = & P_s(PrimU) \\ & + (1 - P_s(PrimU)) P_s(PrimV) \\ & + (1 - P_s(PrimU))(1 - P_s(PrimV)) P_s(PrimW) \\ & + (1 - P_s(PrimU))(1 - P_s(PrimV))(1 - P_s(PrimW)) P_s(PrimX) \\ & + \dots \end{aligned}$$

und damit als Anzahl der zu erwartenden perfekten Basenpaarungen mit irgendeinem der Primer

$$N * P_s(Primers) ;$$

für die Bestimmung von  $P_a(Primers)$  auf dem anti-sense-Strang werden die analogen Gleichungen verwendet; ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge *M* ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet, die Wahrscheinlichkeit dafür ist

$$P_a(Primers) \sum_{i=0}^{M-2} (1 - P_a(Primers))^i ;$$

für große *M* und kleine  $P_a(Primers)$  wird dieses durch folgenden Ausdruck berechnet:

$$\frac{P_a(Primers)}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] ;$$

für die Gesamtzahl *F* der Amplifikate, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit

$$F = N * P_s(Primers) \frac{(P_a(Primers))}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] + N * P_a(Primers) \frac{(P_s(Primers))}{\log(1 - P_s(Primers))} [(1 - P_s(Primers))^M - 1]$$

Formel 1

9. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammt, die in mindestens einer Zelle des jeweiligen Organismus in mRNA transkribiert werden, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
10. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus nach der Transkription in mRNA gespliceten Genomabschnitten (Exons) stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
11. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammen, welche für Teile einer oder mehrerer Genfamilien kodieren, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
12. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8. aus Genomabschnitten stammen, welche für sogenannte „matrix attachment sites“ (MARs)- charakteristische Sequenzen

- enthalten, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
13. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammen, welche als sogenannte „boundary elements“ die Verpackungsdichte des Chromatins organisieren, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
14. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus „multiple drug resistance gene“ (MDR)-Promotoren oder kodierenden Regionen stammen, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
15. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß zur Amplifikation der in Anspruch 1 beschriebenen Fragmente zwei Oligonukleotide oder zwei Klassen von Oligonukleotiden verwendet werden, von denen eines oder eine Klasse außer im Kontext CpG oder CpNpG zwar die Base C enthalten kann, nicht aber die Base G und von denen das andere oder die andere Klasse außer im Kontext CpG oder CpNpG zwar die Base G, nicht aber die Base C enthalten kann.
16. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchgeführt wird, von denen eines eine vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen

würde, wenn ein ebenso langes DNA Fragment, an welches einer der Transkriptionsfaktoren

AhR/Arnt	aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator
Arnt	aryl hydrocarbon receptor nuclear translocator
AML-1a	CBFA2; core-binding factor, runt domain, alpha subunit 2 (acute myeloid leukemia 1; aml1 oncogene)
AP-1	activator protein-1 (AP-1); Synonyme: c-Jun
C/EBP	CCAAT/enhancer binding protein
C/EBPalpha	CCAAT/enhancer binding protein (C/EBP), alpha
C/EBPbeta	CCAAT/enhancer binding protein (C/EBP), beta
CDP	CUTL1; cut (Drosophila)-like 1 (CCAAT displacement protein)
CDP	CUTL1; cut (Drosophila)-like 1 (CCAAT displacement protein)
CDP CR1	complement component (3b/4b) receptor 1
CDP CR3	complement component (3b/4b) receptor 3
CHOP-C/EBPalpha	DDIT; DNA-damage-inducible transcript 3/CCAAT/enhancer binding protein (C/EBP), alpha
c-Myc/Max	avian myelocytomatosis viral oncogene/MYC-ASSOCIATED FACTOR X
CREB	cAMP responsive element binding protein
CRE-BP1	CYCLIC AMP RESPONSE ELEMENT-BINDING PROTEIN 2, CREB2, CREBP1; now ATF2; activating transcription factor 2
CRE-BP1/c-Jun	activator protein-1 (AP-1); Synonyme: c-Jun
CREB	MP responsive element binding protein
E2F	E2F transcription factor (originally identified as a DNA-binding protein essential E1A-dependent activation of the adenovirus E2 promoter)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
Egr-1	early growth response 1
Egr-2	early growth response 2 (Krox-20 (Drosophila) homolog)
ELK-1	ELK1, member of ETS (environmental tobacco smoke) oncogene family
Freac-2	FKHL6; forkhead (Drosophila)-like 6; FORKHEAD-RELATED ACTIVATOR 2; FREAC2
Freac-3	FKHL7; forkhead (Drosophila)-like 7; FORKHEAD-RELATED ACTIVATOR 3; FREAC3
Freac-4	FKHL8; forkhead (Drosophila)-like 8; FORKHEAD-RELATED ACTIVATOR 4; FREAC4
Freac-7	FKHL11; forkhead (Drosophila)-like 9; FORKHEAD-RELATED ACTIVATOR 7; FREAC7



GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-2	GATA-binding protein 2/Enhancer-Binding Protein GATA2
GATA-3	GATA-binding protein 3/Enhancer-Binding Protein GATA3
GATA-X	
HFH-3	FKHL10; forkhead (Drosophila)-like 10; FORKHEAD-RELATED ACTIVATOR 6; FREAC6
HNF-1	TCF1; transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
HNF-4	hepatocyte nuclear factor 4
IRF-1	interferon regulatory factor 1
ISRE	interferon-stimulated response element
Lmo2 complex	LIM domain only 2 (rhombotin-like 1)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
myogenin/NF-1	Myogenin (myogenic factor 4)/Neurofibromin 1; NEUROFIBROMATOSIS, TYPE I
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
NF-E2	NFE2; nuclear factor (erythroid-derived 2), 45kD
NF-kappaB (p50)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p50 subunit
NF-kappaB (p65)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p65 subunit
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NRSF	NEURON RESTRICTIVE SILENCER FACTOR; REST; RE1-silencing transcription factor
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
P300	E1A (adenovirus E1A oncoprotein)-BINDING PROTEIN, 300-KD
P53	tumor protein p53 (Li-Fraumeni syndrome); TP53

Pax-1	paired box gene 1
Pax-3	paired box gene 3 (Waardenburg syndrome 1)
Pax-6	paired box gene 6 (aniridia, keratitis)
Pbx 1b	pre-B-cell leukemia transcription factor
Pbx-1	pre-B-cell leukemia transcription factor 1
RORalpha2	RAR-RELATED ORPHAN RECEPTOR ALPHA; RETINOIC ACID-BINDING RECEPTOR ALPHA
RREB-1	ras responsive element binding protein 1
SP1	simian-virus-40-protein-1
SP1	simian-virus-40-protein-1
SREBP-1	sterol regulatory element binding transcription factor 1
SRF	serum response factor (c-fos serum response element-binding transcription factor)
SRY	sex determining region Y
STAT3	signal transducer and activator of transcription 1, 91kD
Tal-1alpha/E47	T-cell acute lymphocytic leukemia 1/transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
TATA	cellular and viral TATA box elements
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
TCF11/MafG	v-maf musculoaponeurotic fibrosarcoma (avian) oncogene family, protein G
TCF11	Transcription Factor 11; TCF11; NFE2L1; nuclear factor (erythroid-derived 2)-like 1
USF	upstream stimulating factor
Whn	winged-helix nude
X-BP-1	X-box binding protein 1 oder
YY1	ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins

bindet, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

17. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß man die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchführt, von denen eines die eine vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

18. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchgeführt wird, von denen mindestens eine der Sequenzen (von 5' nach 3')

TCGCGTGTA, TACACGCGA, TGTACGCGA, TCGCGTACA,  
TTGCGTGTT, AACACGCAA, GGTACGTAA, TTACGTACC,  
TCGCGTGTT, AACACGCGA, GGTACGCGA, TCGCGTACC,  
TTGCGTGTA, TACACGCAA, TGTACGTAA, TTACGTACA,  
TACGTG, CACGTA, TACGTG, CACGTA,

ATTGCGTGT, ACACGCAAT, GTACGTAAT, ATTACGTAC,  
ATTGCGTGA, TCACGCAAT, TTACGTAAT, ATTACGTAA,  
ATCGCGTGA, TCACGCGAT, TTACGCGAT, ATCGCGTAA,  
ATCGCGTGT, ACACGCGAT, GTACGCGAT, ATCGCGTAC,  
TGTGGT, ACCACA, ATTATA, TATAAT,

TGAGTTAG, CTAACCTCA, TTGATTTA, TAAATCAA,  
TGATTTAG, CTAAATCA, TTGAGTTA, TAACTCAA,

TTTGGT, ACCAAA, ATTAAA, TTTAAT,  
TGTGGA, TCCACA, TTTATA, TATAAA,  
TTTGGA, TCCAAA, TTTAAA, TTTAAA,  
TGTGGT, ACCACA, ATTATA, TATAAT,

ATTAT, ATAAT, GTAAT, ATTAC,  
ATTGT, ACAAT, GTAAT, ATTAC,

GAAAG, CTTTC, TTTTT, AAAAA,  
GTAAT, ATTAC, ATTGT, ACAAT,  
GAAAT, ATTTT, ATTTT, AAAAT,  
GTAAG, CTTAC, TTTGT, AAAAA,  
TTAATAATCGAT, ATCGATTATTAA, ATCGATTATTGG, CCAATAATCGAT,  
ATCGATTA, TAATCGAT, TAATCGAT, ATCGATTA,

ATCGATCGG, CCGATCGAT, TCGATCGAT, ATCGATCGA,  
ATCGATCGT, ACGATCGAT, GCGATCGAT, ATCGATCGC,

TATCGATA, TATCGATA, TATCGGTG, CACCGATA,  
TATTAATA, TATTAATA, TATTGGTG, CACCAATA,

GTGTAATATTT, AAATATTACAC, GGGTATTGTAT, ATACAATACCC,  
GTGTAATTTTT, AAAAATTACAC, GGGGATTGTAT, ATACAATCCCC,  
ATGTAATTTTT, AAAAATTACAT, GGGGATTGTAT, ATACAATCCCC,  
ATGTAATATTT, AAATATTACAT, GGGTATTGTAT, ATACAATACCC,  
ATTACGTGGT, ACCACGTAAT, ATTACGTGGT, ACCACGTAAT,

TGACGTAA, TTACGTCA, TTACGTTA, TAACGTAA,  
TGACGTTA, TAACGTCA, TGACGTTA, TAACGTCA,  
TTACGTAA, TTACGTAA, TTACGTAA, TTACGTAA,  
TGACGTTA, TAACGTCA, TAACGTTA, TAACGTTA,

TGACGT, ACGTCA, GCGTTA, TAACGC,  
TGACGT, ACGTCA, ACGTTA, TAACGT,  
TTTCGCGT, ACGCGAAA, GCGCGAAA, TTTCGCGC,  
TTTGGCGT, ACGCCAAA, GCGTTAAA, TTTAACGC,

TAGGTGTTA, TAACACCTA, TAATATTTG, CAAATATTA,  
TAGGTGTTT, AAACACCTA, GAATATTTG, CAAATATTC,

GTAGGTGG, CCACCTAC, TTATTTGT, ACAAATAA,  
GTAGGTGT, ACACCTAC, ATATTTGT, ACAAATAT,

TGCGTGGGCGG, CCGCCCACGCA, TCGTTTACGTA, TACGTAAACGA,  
TGCGTGGGCGT, ACGCCCACGCA, ACGTTTACGTA, TACGTAAACGT,

TGCGTAGGCGT, ACGCCTACGCA, ACGTTTACGTA, TACGTAAACGT,  
TGCGTAGGCGG, CCGCCTACGCA, TCGTTTACGTA, TACGTAAACGA,  
ATAGGAAGT, ACTTCCTAT, ATTTTTTGT, ACAAAAAAT,

TCGGAAGT, ACTTCCGA, ATTTTCGG, CCGAAAAT,  
TCGGAAGT, ACTTCCGA, GTTTTCGG, CCGAAAAC,  
TCGGAAT, ATTTCCGA, ATTTTCGG, CCGAAAAT,  
TCGGAAT, ATTTCCGA, GTTTTCGG, CCGAAAAC,  
GTAAATAA, TTATTTAC, TTGTTTAT, ATAAACAA,  
GTAAATAAATA, TATTTATTTAC, TGTTTATTTAT, ATAAATAAACA,

AAAGTAAATA, TATTTACTTT, TGTTTATTTT, AAAATAAACA,  
AATGTAAATA, TATTTACATT, TGTTTATATT, AATATAAACA,  
TAAGTAAATA, TATTTACTTA, TGTTTATTTA, TAAATAAACA,  
TATGTAAATA, TATTTACATA, TGTTTATATA, TATATAAACA,

ATAAATA, TATTTAT, TGTTTAT, ATAAACA,  
ATAAATA, TATTTAT, TATTTAT, ATAAATA,  
GATA, TATC, TATT, AATA,

TAGATAA, TTATCTA, TTATTTG, CAAATAA,  
TTGATAA, TTATCAA, TTATTAG, CTAATAA,  
GATAA, TTATC, TTATT, AATAA,

GATG, CATC, TATT, AATA,

GATAG, CTATC, TTATT, AATAA,  
GATAAG, CTTATC, TTTATT, AATAAA,

TGTTTATTTA, TAAATAAACA, TAAATAAATA, TATTTATTTA,  
TGTTTGTTA, TAAACAAACA, TAAATAAATA, TATTTATTTA,  
TATTTATTTA, TAAATAAATA, TAAATAAATA, TATTTATTTA,  
TATTTGTTA, TAAACAAATA, TAAATAAATA, TATTTATTTA,

GTTAATGATT, AATCATTAAC, AATTATTAAT, ATTAATAATT,  
GTTAATTATT, AATAATTAAC, AATAATTAAT, ATTAATTATT,  
GTTAATTAAT, ATTAATTAAC, ATTAATTAAT, ATTAATTAAT,  
GTTAATGAAT, ATTCATTAAC, ATTTATTAAT, ATTAATAAAT,

TAAAGTTTA, TAAACTTTA, TGAATTTTG, CAAAATTCA,  
TAAAGGTTA, TAACCTTTA, TGATTTTTG, CAAAAATCA,

AAAGTGAAATT, AATTCACCTT, GGTTTTATTTT, AAAATAAAACC,  
AAAGCGAAATT, AATTCGCTTT, GGTTTCGTTTT, AAAACGAAACC,

TAGTTTTATTTTTT, AAAAAATAAACTA, GGGAAAGTGAAATTG,  
CAATTCACCTTCCC,  
TAGTTTTATTTTTT, AAAAAATAAACTA, GGAAAAGTGAAATTG,  
CAATTCACCTTTCC,  
TAGTTTTTTTTTTT, AAAAAAAAAAACTA, GGAAAAGAGAAATTG,  
CAATTCCTCTTTCC,  
TAGTTTTTTTTTTT, AAAAAAAAAAACTA, GGGAAAGAGAAATTG,  
CAATTCCTCTTCCC,  
TAGGTG, CACCTA, TATTTG, CAAATA,

TTTTAAAAATAATTTT, AAAATTATTTTTAAAA, AGGGTTATTTTTAGAG,  
CTCTAAAAATAACCCT,  
TTTTAAAAATAATTTT, AAAATTATTTTTAAAA, GGAGTTATTTTTAGAG,  
CTCTAAAAATAACTCC,  
TTTTAAAAATAATTTT, AAAATTATTTTTAAAA, AGAGTTATTTTTAGAG,  
CTCTAAAAATAACTCT,  
TTTTAAAAATAATTTT, AAAATTATTTTTAAAA, GGGGTTATTTTTAGAG,  
CTCTAAAAATAACCCC,

TGTTATTAATAATAGAAA, TTTCTATTTTTAATAACA,  
TTTTTATTTTTAGTAATA, TATTACTAAAAATAAAAA,  
TGTTATTAATAATAGAAAT, ATTCTATTTTTAATAACA,  
GTTTTATTTTTAGTAATA, TATTACTAAAAATAAAAC,  
TTTGGTAT, ATACCAAA, GTGTAAA, TTTAACAC  
GGGA, TCCCC, TTTTT, AAAAA,

TAGGGG, CCCCTA, TTTTTA, TAAAAA,  
GAGGGG, CCCCTC, TTTTTT, AAAAAA,  
TGTTGAGTTAT, ATAACCTAACA, ATGATTTAGTA, TACTAAATCAT,  
TGTTGATTTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAACTCAC,  
TGTTGAGTTAT, ATAACCTAACA, ATGATTTAGTA, TACTAAATCAT,  
TGTTGATTTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAACTCAC,

GGGGATTTTT, AAAAATCCCC, GGGAATTTTT, AAAAATTCCC,  
GGGGATTTTT, AAAAATCCCC, GGGGATTTTT, AAAAATCCCC,  
GGGGATTTTT, AAAAATCCCC, GGAAATTTTT, AAAAATTTC,  
GGGAATTTTT, AAAAATTCCC, GGAAATTTTT, AAAAATTTC,  
GGGAATTTTT, AAAAATTCCC, GGAAATTTTT, AAAAATTTC,  
GGGATTTTT, AAAAATCCC, GGAAAGTTTT, AAAACTTTC,  
GGGAATTTTT, AAAAATTCCC, GGGAATTTTT, AAAAATTCCC,  
GGGATTTTT, AAAAATCCC, GGGAAGTTTT, AAAACTTCCC,  
GGGATTTTTTA, TAAAAAATCCC, TGGAAAGTTTT, AAAACTTTCCA,  
TTTAGTATTACGGATAGAGGT, ACCTCTATCCGTAATACTAAA,  
GTTTTGTTTCGTGGTGTGAA, TTCAACACCACGAACAAAAC,  
TTTAGTATTACGGATAGAGTT, AACTCTATCCGTAATACTAAA,  
GGTTTTGTTTCGTGGTGTGAA, TTCAACACCACGAACAAAACC,  
TTTAGTATTACGGATAGCGTT, AACGCTATCCGTAATACTAAA,  
GGCGTTGTTTCGTGGTGTGAA, TTCAACACCACGAACAACGCC,  
TTTAGTATTACGGATAGCGGT, ACCGCTATCCGTAATACTAAA,  
GTCGTTGTTTCGTGGTGTGAA, TTCAACACCACGAACAACGAC,

ATATGTAAAT, ATTTACATAT, ATTTGTATAT, ATATACAAAT,  
TTATGTAAAT, ATTTACATAA, ATTTGTATAA, TTATACAAAT,

GAATATTTA, TAAATATTC, TGAATATTT, AAATATTCA,  
GAATATGTA, TACATATTC, TGTATATTT, AAATATACA,

ATAAT, ATTAT, ATTAT, ATAAT,  
GTAAT, ATTAC, ATTAT, ATAAT,

AATGTAAAT, ATTTACATT, ATTTGTATT, AATACAAAT,

ATTTGTATATT, AATATACAAAT, GGTATGTAAAT, ATTTACATACC,  
ATTTGTATATT, AATATACAAAT, AATATGTAAAT, ATTTACATATT,  
ATTTGTATATT, AATATACAAAT, AGTATGTAAAT, ATTTACATACT,  
ATTTGTATATT, AATATACAAAT, GATATGTAAAT, ATTTACATATC,

AGGAGT, ACTCCT, ATTTTT, AAAAAT,  
GGGAGT, ACTCCC, ATTTTT, AAAAAT,  
GGATATGTTTCGGGTATGTTT, AAACATACCCGAACATATCC,  
GGATATGTTTCGGGTATGTTT, AAACATACCCGAACATATCC,  
GGATATGTTTCGGGTATGTTT, AAACATACCCGAACATATCC,  
AGATATGTTTCGGGTATGTTT, AAACATACCCGAACATATCT,  
TCGTTTCGTTTTAGATAT, ATATCTAAAACGAAACGA,  
ATATTTAGAGCGGAACGG, CCGTTCCGCTCTAAATAT,

CGTTACGGTT, AACCGTAACG, AATCGTGACG, CGTCACGATT,  
CGTTACGGTT, AACCGTAACG, GATCGTGACG, CGTCACGATC,  
CGTTACGTTT, AACCGTAACG, AAGCGTGACG, CGTCACGCTT,  
CGTTACGTTT, AACCGTAACG, GAGCGTGACG, CGTCACGCTC,



TTTACGTATGA, TCATACGTAAA, TTATGCGTGAA, TTCACGCATAA,  
TTTACGTTTGA, TCAAACGTAAA, TTAAGCGTGAA, TTCACGCTTAA,  
TTTACGTTTTA, TAAAACGTAAA, TGAAGCGTGAA, TTCACGCTTCA,  
TTTACGTATTA, TAATACGTAAA, TGATGCGTGAA, TTCACGCATCA,

AATTAATTAA, TTAATTAATT, TTGATTGATT, AATCAATCAA,  
TATTAATTAA, TTAATTAATA, TTGATTGATG, CATCAATCAA,

TAATTAT, ATAATTA, ATGATTG, CAATCAT,

TAGGTTA, TAACCTA, TGATTTA, TAAATCA,

TTTTAAATATTTTT, AAAAATATTTAAAA, GGGGGTGTTTGGGG,  
CCCCAAACACCCCC,  
TTTTAAATTATTTTT, AAAATAATTTAAAA, GGGGTGGTTTGGGG,  
CCCCAAACCACCCC,  
TTTTAAATTTTTTTT, AAAAAAATTTAAAA, GGGGGGGTTTGGGG,  
CCCCAAACCCCCCCC,  
TTTTAAATAATTTTT, AAAATTATTTAAAA, GGGGTTGTTTGGGG,  
CCCCAAACAACCCC,

GAGGCGGGG, CCCC GCCTC, TTTCGTTTT, AAAACGAAA,  
GAGGTAGGG, CCCTACCTC, TTTTGTTTT, AAAACAAAA,  
AAGGCGGGG, CCCC GCCTT, TTTCGTTTT, AAAACGAAA,  
AAGGTAGGG, CCCTACCTT, TTTTGTTTT, AAAACAAAA,

GGGGGCGGGGT, ACCCCGCCCCC, ATTTTCGTTTTT, AAAACGAAAT,  
GGGGGCGGGGT, ACCCCGCCCCC, GTTTTCGTTTTT, AAAACGAAAC,  
TATTATTTTAT, ATAAAATAATA, GTGGGGTGATA, TATCACCCCAC,  
GATTATTTTAT, ATAAAATAATC, GTGGGGTGATT, AATCACCCCAC,

ATTACGTGAT, ATCACGTAAT, ATTACGTGAT, ATCACGTAAT,  
ATTACGTGAT, ATCACGTAAT, GTTACGTGAT, ATCACGTAAC,

TTTTATATGG, CCATATAAAA, TTATATAAGG, CCTTATATAA,  
TTATATATGG, CCATATATAA, TTATATATGG, CCATATATAA,  
AAATAAT, ATTATTT, GTTGTTT, AAACAAC,  
AAATTAA, TTAATTT, TTAGTTT, AAATAA,  
AAATTAT, ATAATTT, GTAGTTT, AAATAA,  
AAATAAA, TTTATTT, TTTGTTT, AAATAA,

ATTTTTCGGAAATG, CATTTCCGAAAAAT, TATTTTCGGGAAAT,  
ATTTCCCGAAAATA,  
ATTTTTCGGAAATG, CATTTCCGAAAAAT, TATTTTCGGGAAAT,  
ATTTCCCGAAAATA,  
ATTTTTCGGGAAATG, CATTTCCCGAAAAT, TATTTTTCGGAAAT,  
ATTTCCCGAAAATA,  
ATTTTTCGGGAAGTG, CACTTCCCGAAAAT, TATTTTTCGGAAAT,

ATTTCCGAAAAATA,

AATAGATGTT, AACATCTATT, AATATTTGTT, AACAAATATT,  
AATAGATGGT, ACCATCTATT, ATTATTTGTT, AACAAATAAT,

GTATAAATA, TATTTATAC, TATTTATAT, ATATAAATA,  
GTATAAATG, CATTTATAC, TATTTATAT, ATATAAATA,  
GTATAAAAA, TTTTATAC, TTTTATAT, ATATAAAAA,  
GTATAAAAG, CTTTATAC, TTTTATAT, ATATAAAAA,  
TTATAAATA, TATTTATAA, TATTTATAG, CTATAAATA,  
TTATAAATG, CATTTATAA, TATTTATAG, CTATAAATA,  
TTATAAAAA, TTTTATAA, TTTTATAG, CTATAAAAA,  
TTATAAAAG, CTTTATAA, TTTTATAG, CTATAAAAA,  
GGGGGTTGACGTA, TACGTCAACCCCC, TCGGTTAATTTTT,  
AAAAATTAACGCA,  
GGGGGTTGACGTA, TACGTCAACCCCC, TACGTTAATTTTT,  
AAAAATTAACGTA,

TGACGTATATTTTT, AAAAATATACGTCA, GGGGATATGCGTTA,  
TAACGCATATCCCC,  
TGACGTATATTTTT, AAAAATATACGTCA, GGGGGTATGCGTTA,  
TAACGCATACCCCC,  
ATGATTTAGTA, TACTAAATCAT, TGTTGAGTTAT, ATAACCTAACA,  
GTTAT, ATAAC, ATGAT, ATCAT,

TTACGTGA, TCACGTAA, TTACGTGG, CCACGTAA,  
TTACGTGG, CCACGTAA, TTACGTGG, CCACGTAA,  
TTACGTGG, CCACGTAA, TTACGTGA, TCACGTAA,  
TTACGTGA, TCACGTAA, TTACGTGA, TCACGTAA,  
GACGTT, AACGTC, AGCGTT, AACGCT,

TGACGTGT, ACACGTCA, ATACGTTA, TAACGTAT,  
TGACGTGG, CCACGTCA, TTACGTTA, TAACGTAA,  
CGGTTATTTTG, CAAAATAACCG, TAAGATGGTCG oder CGACCATCTTA

enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

19. Verfahren nach einem der Ansprüche 16 bis 18, dadurch gekennzeichnet, daß die zur Amplifikation verwendeten Oligonukleotide außer den in den Ansprüchen 16 bis 18 definierten Konsensussequenzen mehrere Positionen enthalten, an denen

entweder irgendeine der drei Basen G, A und T oder irgendeine der Basen C, A und T vorhanden sein kann.

20. Verfahren nach Anspruch 19, dadurch gekennzeichnet, daß die zur Amplifikation verwendeten Oligonukleotide außer einer der in den Anspruch 18 beschriebenen Konsensussequenzen nur maximal zusätzlich so viele weitere Basen enthalten, wie es zur gleichzeitigen Amplifikation von mehr als einhundert verschiedenen Fragmenten pro Reaktion aus der chemisch behandelten DNA, berechnet entsprechend Anspruch 8, erforderlich ist.
21. Verfahren nach einem der vorangehenden Ansprüche dadurch gekennzeichnet, daß die Untersuchung des Sequenzkontextes aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide gemäß Anspruch 1 c) durch Hybridisierung der bereits in der Amplifikation mit einem Fluoreszenzmarker versehenen Fragmente an einen Oligonukleotid- Array (DNA Chip) erfolgt.
22. Verfahren nach einem der Ansprüche 1 bis 20 dadurch gekennzeichnet, daß die amplifizierten Fragmente auf einer Oberfläche immobilisiert und anschließend eine Hybridisierung mit einer kombinatorischen Bibliothek von unterscheidbaren Oligonukleotid- oder PNA-Oligomer-Sonden durchgeführt wird.
23. Verfahren nach Anspruch 22 dadurch gekennzeichnet, daß die Sonden mittels Matrix-assistierter Laser-Desorptions/Ionisations Massenspektrometrie (MALDI-MS) anhand ihrer eindeutigen Masse nachgewiesen werden und damit der Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide entschlüsselt wird.
24. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß die Amplifikation wie beschrieben in Schritt b) des Anspruchs 1 durch eine Polymerase Kettenreaktion ausgeführt wird, in der die Größe der amplifizierten Fragmente mittels auf weniger als 30 s verkürzter Kettenverlängerungsschritte

begrenzt wird.

25. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß nach der Amplifikation gemäß Schritt b) des Anspruchs 1 die Produkte durch Gelelektrophorese aufgetrennt werden und die Fragmente, die kleiner als 2000 Basenpaare oder kleiner als ein beliebiger Grenzwert unterhalb von 2000 Basenpaaren sind, durch Ausschneiden von den anderen Produkten der Amplifikation vor der Auswertung gemäß Schritt c) des Anspruchs 1 abgetrennt werden.
26. Verfahren nach Anspruch 25 dadurch gekennzeichnet, daß nach der Abtrennung von Amplifikaten bestimmter Größe diese vor der Durchführung des Schrittes c) des Anspruchs 1 nochmals amplifiziert werden.
27. Kit, enthaltend mindestens zwei Primerpaare, Reagenzien und Hilfsstoffe für die Amplifikation und/oder Reagenzien und Hilfsmittel für die chemische Behandlung gemäß Anspruch 1 a) und/oder eine kombinatorische Sondenbibliothek und/oder einen Oligonukleotid-Array (DNA-Chip), soweit sie für die Durchführung des erfindungsgemäßen Verfahrens erforderlich oder dienlich sind.

## SEQUENZPROTOKOLL

## ALLGEMEINE ANGABEN:

## ANMELDER:

NAME: Epigenomics AG  
STRASSE: Kastanienallee 24  
LAND: Berlin  
POSTLEITZAHL: 10435  
TELEFON: 030-243450  
TELEFAX: 030-24345555

BEZEICHNUNG DER ERFINDUNG: Verfahren zur parallelen Detektion  
des Methylierungszustandes von  
genomischer DNA

ANZAHL DER SEQUENZEN: 31

## COMPUTERLESBARE FASSUNG:

DATENTRÄGER: Diskette  
COMPUTER: IBM PC-kompatibel  
BETRIEBSSYSTEM: PC-DOS/MS-DOS

## DATEN DER JETZIGEN ANMELDUNG:

ANMELDENUMMER: nicht bekannt  
ANMELDETAG: 6.12.2000

## ANGABEN ZU SEQ ID-NO:1:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:1:

AGTAGTAGTA GT

12

## ANGABEN ZU SEQ ID-NO:2:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:2:

AAAACAAAAA CC

12

## ANGABEN ZU SEQ ID-NO:3:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 973 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:3:

AGTAGTAGTA	GTAGCGTTTT	TGAAGTTTTT	TGTGGAAGGT	GAGAAATTTA	TCGATAAGTT	60
TTTATAGTAGG	AGTGTTTTTG	GGGAGGGAGT	GAGTGGGAGA	TTAGAAATGG	GGTCGGTGGA	120
ATATTTTTGT	AAAATTTTAG	GAATTATTAG	TATTTTATTT	TTTTTTATAA	AGATTGTTTT	180
TATGTTTGAG	TTGTTTTATA	TAGTGTAGAA	ATTGAGAATG	TAAACGTGTA	AACGATCGGT	240
GTGATTTATT	GAAAGGCGAT	GCGCGTTTTT	TTTTTGTAGG	TAGAATTGTT	TTAGGAAGTA	300
TGAAAGTGGA	CGTAGTTCGG	TGTTAGGTGT	TAGCGATTTG	AGTTCGTTTT	CGGGTTTTAT	360
TTATTTTTAG	TTTCGGTTTT	AGATATTTTT	CGAGGCGTTT	TTTTTTGTTC	GTTCGATTTT	420
TGAGCGGAGC	GTTTCGGGGT	GTGAGGAGAA	TCGGTAAATT	TTTCGGGGCG	TTGGGCGTCG	480
TAGAGTCGTC	GCGCGTTTAG	TTTCGTTACG	TTGGTTGTCG	AGCGTATTTT	GGTTGCGTTC	540
GGCGGGGAGG	CGTCGAGGGT	AGTTAAGGGG	AGTAGGTTAC	GTGAGAGGAG	GAGTTTGATT	600
TATTTTTTAG	GCGGTAGGCG	TATGCGTATT	TTTTATTGTC	GTTTCGGTCG	GGAGGTTTAC	660
GTGGAATCGT	ACGTGTTTGG	TTTTGTAGTT	AGGGGTTTTT	GGTTCGGGGC	GCGTAGGGGC	720
GGGTTCGTAG	TGGGATTCGC	GGAGAGGGGC	GCGGCGGGGC	GGAGCGTTTG	GAGATTTAGT	780
TTGCGGGTTT	CGTAATTATT	ATTCGTGATA	ATTATAGTTT	TTGGAGAGTT	GTTTAGGTTT	840
TTTGCGGGGT	TTTTTACGAA	TTTATTATAT	TTAATATTCG	TGAATTTAGA	AATTAAGATA	900
AAGATTGTAT	TTTTGTTTTG	TAAATTATTA	GTTTCGTAAT	TTAGAATTTA	TTTTGTAAAT	960
GGGTTTTTGT	TTT					973

ANGABEN ZU SEQ ID-NO:4:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1890 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:4:

AGTAGTAGTA	GTAAAGTTTA	AGAGTGATAT	TTTTTTAGAT	ATAGTATAAT	ATTGTTTTTG	60
TAAAAGGTTA	TTTATATGAA	TATAATGTTT	TTTGGGTTAA	TTAATTCGTT	TTGAGAATAG	120
TTAGGTTAAA	ATTTTTAGGT	TTTTTATTTT	GGGTTATTTA	AGTTAGGATC	GCGACGTGAG	180
TTTCGGGGTGA	GTCGAGGGGT	ATTCGGGGTC	GAGGTTATTC	GAATCGAAGT	TATGAGGCGC	240
GGGTCGGTAA	TCGGAAGTGG	TTTAGGGAGA	GTTGTACGAG	ATTCGGGGGT	TGTGATTTGG	300
AAATAAAATA	AAATAAAATA	AAATTTTAAT	TGTTTTGGGA	TTGATTTTTA	AAAGAGCGTT	360
TTTTTTGGTTT	AAGAGGTCGG	CGTTCGCGGA	GATGTCGTTT	TAAAGGGTTG	TTTTTTAAGC	420
GTGTAGGTCG	CGTACGGGTT	TTTTTAGCGG	GCGGGTAAAA	TGGGCGTCGG	TATTCGGGAG	480
GCGTTTGTTT	AGGCGTTCGG	GCGTCGTTTA	TAGAGTACGT	TCGTTTGCGG	TTTTAGAGCG	540
TCGTTTTTTT	GTCGTTTTTC	TCGTTTCGGT	TGACGTTTCG	ATCGCGGTCG	GTTATCGTTT	600
TTCGTTTCGA	CGGTTACGTT	TGTTTTAAAT	CGCGCGGGCG	TTTTTTAGGT	GTTGTTGGGC	660
GGGTTTCGTT	TTTCGTGTTT	AAGGTTTCGT	TTTCGCGCGT	AGTCGCGCGT	TCGTTGTTTT	720
TTTTTCGTTT	TTATAGTTTC	GTTTTTATAG	TTTCGTTTCG	TTTTTAAGTT	TCGTTTTTTA	780
GGAATTCGCG	CGTCGAAGGT	TAGGTTTGGG	CGGAGCGTAT	AGCGTTGGGC	GTTGGGGAGG	840
TTGCGTCGTA	GTATTCGGTT	GGTTAGGATT	AAGTGGGTTT	GAGGCGGACG	TGAGAAGGGT	900
CGGGTTAAGA	TGGCGGTGTA	GGTGGTGTAG	GCGGTGTAGG	CGGTTTATTT	CGAGTTTGAC	960
GTTTTTTTCG	TTTGTTTTAA	TTACGTTTTG	AGTATAGAGA	AGGAGGAAGT	AATGGGGTTG	1020
TGTATAGGGG	AGGTGAGTAG	GTTTGTTAGT	TTGGATGGAA	TTTTGTTGAG	TAGTTTTAGT	1080
GTGTTTTTCG	GGTGGGTGTC	GGTAGTTTTT	AGGGTTGCGG	AGGTTATAGG	TATTTTCGAT	1140
TTAGGTTTTT	GGATATTTTT	TATATAGTTT	TGTCGCGGGG	AGTTTCGTTT	TTTTTGGTCT	1200
TTTGATATTT	GTTTAGTTTT	TTGCGAGTTT	TCGGCGGTTT	GTATAGGTTT	CGGGAGTTTT	1260
GTTTTTTTTT	TTGTAGTTTG	GGGTTTTTTT	TTAGATTTGC	GTTTTTCGTT	CGGTTGATTA	1320
AAATTGTAAG	GTAGGTTAGA	AAGATATTGG	AGTATAATGA	GGATGTTTCG	GTATTACGAC	1380
GTAGTAGTTG	ATTATAGTTA	GAGTTTTTTG	TTTCGTTTTC	AGTTTTTTGT	TTAGGGAGTA	1440
GAGATATTAA	TTAAAGTATT	TGAAGGGTAT	CGAAGAGTTT	AATAGAAGGT	GCGGGGTTTG	1500
AAGGAAGTAA	AAGTTTTTCG	TGTATTGTGT	TGAGGAGGGG	TCGAAGAGGA	TGAGGAAATA	1560
TAGTTTAGTT	GTTTATAGTT	TAAAGTAATT	TTTTAGTTTT	TTATATTATG	TGCGTGAATA	1620
TATGATTTAA	TTGTTATATA	ATTTGTATTT	ATATATGTTA	AATAAACGTA	ATGTGATTAA	1680



TAATGTTTTT	TGTTTTTTTT	TATTTTAGTT	GAACGATGAT	ATAAGGTAAG	ATTGTATTTG	1740
TTTATTTATA	TTTTATAATT	TTTAAGTTAT	TGTTAATTTA	AAATTATGGC	GATTTTGGTA	1800
GTATTGTGTT	CGGATCGTGT	TTTAAATTAA	TTATTATTAT	TATTTTCGGG	TTTTTTTAAT	1860
GATATAATTT	TGAATTTTGG	TTTTTGTTTT				1890

ANGABEN ZU SEQ ID-NO:5:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 2222 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:5:

AGTAGTAGTA	GTTTTGGAGA	GTAAGTGGGT	TTTATGTAAA	GTGTTATTTG	TTATATATTG	60
TTATTTTAN	AGGATTTTAG	AATATTTTCG	AATTCGAAAA	TAGAGTGAGT	GTGGAGGGGA	120
GGGGGGTGTT	TACGTGGAGG	AGGGTCGCGA	GAAGGAGAGT	GCGTCGGGTT	GATGGTTAGC	180
GGTGTTCGG	GGGTCGATGT	CGGGCGGGAT	CGGCGGCGCG	GGGGTGGGGC	GACGGCGGGG	240
CGGCGGTCTGA	GTAAGTAGAGG	GGCGTTTCGT	AGAGGTCGGG	GGGGGCGCGG	GGTTTCGCGT	300
CGTTTTGTTA	GTAAGTCGTTT	CGTCGTTTTT	TTCGGTTCGT	TCGTAATCGT	CGCGGGATTC	360
GTCGGTTCGG	GTTTTTGAA	TTTTTTACGG	GGGATAGTCG	ATCGGGGGCG	TCGGGGTTTCG	420
TTTTCGTATT	CGATTCGGAC	GGCGGTAGTA	GGGGGAGGGA	TGGTGTATCG	AGGAAGTTAA	480
GGTTTTATTA	TTTTGTTCGG	CGGGTCGGCG	GTTTCGTTGGT	TCGTTGGTTC	GGAGAAGTGT	540
TGCGTTTAGG	TTGGTTCGTA	GGAAACGGCG	GCGGCGGTTT	ANTTNTANTT	TTNNNNNTNT	600
TNNTTGNNAN	TNTTTTTTTA	TAGGGGAGGG	GTAAGCGGTT	CGCGGGTTTC	GCGTCGGTTCG	660
GGCGATTGGG	TACGCGAGGG	AGCGGGTAGG	GTAAGGGGAA	ATAAATTAAG	GTCGAGATTT	720
AGAGTCGGAG	AGCGCGGGAG	GAGTAGCGGC	GAGAGGTAGG	AGAGGTAGAG	AAGAAGAAAG	780
GAGGGAGAGA	GGGGGCGAAG	ACGGTCGGGA	AATTCGGGGT	TGTAGTTTCG	CGTCGCGTCG	840
GAGTCGTGAC	GGATTTATTG	TAGCGGTCGC	GTTGATTTCG	AATTTTAAGG	TACGAAAAAA	900
GGGGAGGGTA	GAGAGGGAGG	GGAAAGCGGA	GTGTTGTAGC	GTCGGGGCGG	GGGCGGGTTT	960
TCGCGAGCGT	CGTATATTTCG	GGAATTTGTN	GTTTCGTTGC	GGGCGAGTGT	CGCGTGTTTT	1020
GGCGAGTTTT	GGTTGGGGAA	NTTTGAGCGC	GCGGATTACG	TTCGTTTTTT	AGTCGGTTCG	1080
TTTTTGCGTG	TAGGGTGGGG	GGAGGTTAGG	GGCGGGGGCG	CGGACGTCGG	TGACGTCGCG	1140
TGGCGGAGTT	TTTTCGGTTA	TGTGGTTGGA	GGCGGGCGGG	GAAGAAGGTA	AGGTTGGGAG	1200
GGGAAGGAGG	AAATGCGAGG	GTTTTTCGGC	GGGAGGAAGC	GCGAGGGGTG	CGCGAGTGTT	1260
AGCGAAAGGG	GGATAGGGGT	TTTGGAGTGG	AAGTTTTGCG	GGAGAGGAGG	AGTAGAGGAG	1320
TAGTTTTTAGG	GTGTTAGAAT	TATTCGGATG	TCGTATTAAA	AAATAGGAAA	AATTTAAGTT	1380
TTGCGAGAGG	TAGTTAAAGG	TATTTGTGTA	TTCGTGTATC	GATTTGGATT	TATTATAGAA	1440
TTGTACGGTT	TATTAGGATT	GTTGTTATTT	CGGTGTAGTA	TTCGTAGGTT	TATTTGTAAT	1500
ATGAATTGGT	AGAGTTAGGA	ATTAGGTTGT	AAAGATAGAG	AAGCGAGTTA	AAGTTGGTCG	1560
TAGTTCGAGG	CGTGGGGAGA	ATTTGGGTAA	ACGAGGAGAA	GGGATATTTT	TTATTCGTAAG	1620
AAAGATTTTT	TATTTAGTTT	TGTATTTTTA	TAAATCGTTT	AGATTTTTTT	TTGGCGGAGT	1680
TTATTAGTTT	TTGTTTAAAA	AAAGAAAAAA	ATTCGAATGT	AGATTGTTTCG	TTTATTTTTT	1740
TAGTAGAGTT	ATATTTATTT	TTGTGGTAAT	TATTTTTTTA	GAAAATTTAG	ATTATAATAG	1800
GAAATTATAT	TTAGAAAAGT	ATAAGGAGGA	AATTTATGTT	CGAGAGAAGA	AATAAAGTTG	1860
TTAGGAGAGG	TGTGATGAGG	ATAACGAAGA	AAATATTTTG	TAATTATTTT	AATTAGTTAT	1920
TTTTTTATAA	GTTTTGATAA	TCGTTTCGTAT	TCGTGTGTGT	TTGAGAGTGT	GTGTGTGTGT	1980
GTGTGTGTGT	GTGAGAGAGA	GAGAGAGGAT	GTTTTTTGAG	TGATCGGAAA	TTTTTTATTG	2040
ATAAAGTTTT	TGAGTTATTT	TATAAAAAGT	TATTGTGTTG	TTTTTTTTTA	AAAAAGTTTT	2100
AATGATTTTT	ATATGAGAAT	TTTATTTTTT	TTATGTGTTA	ATTATTTTTT	TATATGATGA	2160
GTATGTATTT	TTTTTTAATA	AACGTTTAGC	GATTTGTTTT	AGGATTAAAT	GGTTTTTGTT	2220
TT						2222

ANGABEN ZU SEQ ID-NO:6:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 307 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:6:

AGTAGTAGTA	GTTTCGTTAC	GAATACGATT	AGTTTATATT	GGTTGTTTAG	TATTTTCGGTT	60
AGTTTTTGTG	GGACGTTGGC	GCGTATATCG	TTTCGTCGGG	GGGTCGGGGA	TTTCGGGGAT	120
TTCGGGGTCG	AGGACGAGGG	AGGCGAGTAG	GTCGTTGGTT	TCGACGATTT	TGCGATCGAG	180
GCGTTTGTAG	TTGTTTCGGG	CGAGGCGGGC	GATTTTAGGT	TTTCGAGTAT	TTTCGAGATA	240
TTTCGCGTTAA	GTTTGGTTTC	GGGGGTGGTT	GTCGTCGGGT	TTTCGTTTGC	TTTTTGGTTT	300
TTGTTTT						307

ANGABEN ZU SEQ ID-NO:7:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 523 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:7:

AGTAGTAGTA	GTTGTTGTTG	TTTCGAGGTTT	CGGTGTAGTT	GGAAGTTTTG	GAGGTGGGGA	60
GAGGGATGTT	AGGTAAGTGG	TACGGCGAGC	GTAAGGGAAG	GGGTTAGTTA	TTGATTAGCG	120
GTAGTAATTG	TAGGAATCGT	CGTCGTAGTT	GTAGTCGTTT	TTTCGTTTCG	TTTTTCGGGT	180
TTTCGGGAAA	ATGGTTGTGG	GGTTGGTTCG	GTCGTTAAGT	TTGTTTTTCG	GCGGTGAAGA	240
GCGGGTTGTT	TGGGGGAGTC	GTTTTTTAAT	TTTCGCGCGC	GTTTATTTTT	GCGGAGTTCG	300
TGGTAGGATT	CGAGGGGTTA	CGAGTTGATA	TTTTTTTGGG	TTGTATAAAA	AGTTGAGGCG	360
GGCGTTGGGA	GGAGGTAGCG	GTTGTTGCGG	TGCGGTTTTT	TTTTTTTTTT	ATATTTTGTT	420
CGGGTTATTT	TTTTTTTTTT	TTTTTTTCGT	TTTCGTTTTT	TTTTTTTACG	CGGGTTTTTC	480
GGGGTTTTTG	CGGTTTTTCG	TTGAAGGTCG	CGGTTTTTGT	TTT		523

ANGABEN ZU SEQ ID-NO:8:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 653 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:8:

AGTAGTAGTA	GTAGCGTCGG	GAGTTCGCGT	AGAAACGATT	TGATTTTTTC	CGGCGTTATT	60
TTTTTTTTCG	ATTGGCGTTG	CGCGGAGAAA	ATTTAGTTTG	TCGATGTTTC	GTTTATTATT	120
TCGTTTTTAT	TTATCGCGCG	GTTTTTTGGG	AATTGTAGTT	TTTTCGTGAG	CGGTCGGTAG	180
CGGCGAGGTC	GTCGGCGGAT	TTTAGCGTTT	AGGAAGTTTT	GCGTAGGCGT	AAGCGTTTTT	240
TAGGTAATTA	CGTCGCGGTT	CGAGGTTGTA	GATTTTGGTA	GGGGATATAA	TTGGAGAATG	300
AGGGTGGTCG	GGTGGGAGTA	GGGATTAGGT	TTTATTTAGA	GTTTCGGAGT	TGGAGTTCGT	360
ACGATTTCGAG	GTTTTTTTTT	TTTTTTGTTT	TATAGAAAGA	GCGGAAAGTT	TAAGAATCGG	420
GTCGCGTTTG	GTTGAGTTTG	ATAATGTTTC	GGTTTTACGC	GTATTGACGT	CGATTTTTGT	480
ATGTTTAGTG	TCGTTGCGGG	GTTGGTATTG	CGGTTTCGGG	TTGGCGTTGA	GGAGTTTGGT	540
TTAGTTCGTT	TTTGTTTTTC	GCGGAGCGTT	TGGTCGTGGG	TAAGTTTAGG	TTAGGTTGTT	600
TCGGGGACGT	AGGGTCGCGT	AGACGTTTTT	TTACGTTTCG	GGGTTTTTGT	TTT	653

ANGABEN ZU SEQ ID-NO:9:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1461 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:9:

AGTAGTAGTA	GTAGGCGGTG	GTTAGGTCGG	GTCGGGCGGT	TATTAGCGGT	TTTCGGAGGT	60
ATTTAGGTTT	TACGGTCGAG	TTTTTGGTCG	GAGATCGGTC	GTTTACGGGG	GGTTATAGTA	120
GCGGTCGTTT	GTTTAGGATG	GAGAGGCGGG	TTTTAGGTTT	GGTTCGGAGC	GAGTTTTTTA	180
GGGTTTGTCT	ATACGGCGGG	GTTTCGGTGGT	CGGTATTTGG	TTCGTACGTG	TTCGAGGGGT	240
TTTCGGGTTT	TCGGTATTAT	GGTTATTATC	GGGGTTTCGA	TTACGACGAG	GTCGATGGTT	300
CGGGTAGCGG	GGGCGGCGAG	GAGGTTATGG	TCGGGGTTTA	CGACGCGTTA	TTTTTCGTAC	360
GATACGCGTT	TTCCGGGCGT	ATCCGGGCGT	CGTTTAGGAT	TTTTCCGGGT	TCGGGTTCCG	420
TTTGCGTTTC	GTTTTTTCGG	TACGGTCCGC	GATTTTTTAA	CGGTTATTAT	TCGGCGTACG	480
GATTGGTTAG	GTTTCGCGGG	TCGGGTTTAA	GGAAGGGTTT	GTACGAATTT	TATAGCGAGA	540
GTGACGATGA	TTGGTGTTAA	GTTCCGGGCGA	GGTGGCGTTC	GTTCCGGTTT	TTACGTATTT	600
TACGTATATA	TTTTATTCTA	GGAGTCGCGT	AGAGGTCGCG	GGGGTTTAGT	ATAGAGGGTT	660
CGGGAGAGGG	TTAGTCGGGA	GATTTTAGAT	TTTGGAGAGG	TTAGGGTTGG	GTTATAAGGG	720
TGTTTCGTAG	AGATTTTCGG	TTAAAAGAGA	TTTTTTTGGG	TAGTTACGGC	GTTTTTTAAT	780
TAGTTTCGAT	TTTTTTATTT	ACGATAGGGG	TTTTCGGGTG	GGAGGTAGGG	AGTAGATAAA	840
TTATATAGTT	AAGGGATTGT	AATTAATTTA	GTTATTTTGT	GAGAATTTTG	GGGAATATGA	900
AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	TTTTTAAAG	AAAAAACGGG	GAGAAAAAAA	960
TAGTTTTTAT	TGATGAGTTT	TATTATTTTA	ATTGAATTTT	TTTTTTTTTT	GATGAAGATA	1020
GTTGGTGGTC	GAGTGCGGTA	AAGAAGTTAG	AAGGAATTAG	AATTTTAGTG	TTTTATATTT	1080
ATTATTAGAT	ATATTTATAT	TTATATACGT	TTTLAGATAT	ATATAAGAGT	GTTTGTCCGT	1140
TATATTAAAT	TTTATTATTA	TTGTTTGTAG	AAATTAATTT	AAAAAAATAA	TAATAATAAT	1200
AAATAATTTT	AAAAAGGATA	AAAAAATTAA	TGATTGAGAA	AAGAGGTATT	TTTTTTTGAT	1260
ATTTGGTTTT	GTTTGAAATA	ATAAAAGAAG	AAGAAAAATT	TATTATTATT	ATCGATTTTT	1320
TTGTTTTTTT	TTTTTTTTTT	TTTTATTTTG	TTTGAAAATC	GTGGGTTTGG	GATTGTGAAT	1380
TATTGTATGA	TATTTAAAAA	GAAAAAATAA	ATAAAAAATA	GTTGAATTAA	AGGGTTTTTG	1440
GATAGGAGTG	GTTTTTGTTT	T				1461

ANGABEN ZU SEQ ID-NO:10:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 2536 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:10:

AGTAGTAGTA	GTAATCGGGT	CGTCGTCGTC	GCGGGGTAA	GAGATTACGT	TTAGTAAAAA	60
TTTTTTTAGG	AAAGAGGTTT	GAGAGTTTTT	TTTTTATTTT	GTAAAGGGTG	TTGATTTTGT	120
CGCGTTACGT	TACGTATTAA	AAGGTTCCGC	AAGAAAAACG	CGGGCGAAGT	TTAGTATTTT	180
CGAATAGAAA	GGGAATTTAT	TTTAGAAGGG	GGGTTTACGG	GTTTGAGGGT	TTTTGGGTGG	240
GGAAAGTAGT	ATTGGAGGGA	GAAGATAGAA	GGGATGGTAA	GATAGAGGAA	ATTACGGGGA	300
TAAAAAGGAA	GGGATGAGGA	GTTTTTTTAT	TATAAAAGGG	ATTGACGGAC	GGGATGTAGA	360
GTAATGGGTT	TTTTTCGAAA	AATGGATTGG	TAGGTTTTTT	CGTATTTTTT	AATAATTTCT	420
TATTATTTGA	AAGCGGGGAG	TTTGAGTTTG	GGGTTGCGTA	GATTAAAGTA	GGGTAAGACG	480
AAGGGAAAGA	GGGAGGGGAG	TTTGGAAGTA	AAGTGAATAA	AGAGGTAAAA	GAGGTAAAGT	540
TGGAAATGGA	AACGAGAAAG	GGAGTTGGAA	AGGGGTCGTT	TTGGAAGGGT	TTTGAAAAAG	600
GTTTAGAAAT	GGGATGTTGG	GGAGGTAAAG	GGGGATAGTT	TCGAGGAAGG	GTATTTTAGG	660
AATGTAAGGG	AGCGTTTATG	GTTTTGGGAA	GAGGAAGGGA	TGGGAGGGCG	GGGAGGTAGT	720
CGAGGTTCCG	GAGTTGGAAT	AGAGGGGGTT	TTGCGCGGGG	TTTAGGAGGG	GAGAGTAGGG	780
TCGGAGGGAG	GCGATTATCG	CGGGAAGAGT	TAGAGGATGG	GAGGGAGGGG	TGTTGCGGTT	840
GAAGAGGGGA	TGTGGATTTT	TTTAATGTTA	GGGAGAAGAG	GGAGGTGAAG	CGAGATAAAA	900

GGAAAGGGGC	GAATAGGGAG	AAGAGGAAGG	AGTTTGGCGAC	GCGATGGGGG	AGGGGAGGGA	960
GGTCGTGGGA	AGCGGTCGGG	GGGCGCGGGG	ATGGGAAGGG	GTCGGGCGGG	CGGCGTGGTT	1020
ATTTTAGGTT	GCGGATTTTT	TTTTTGGGGA	GCGGCGTTGT	CGGCGGGCGG	GTTTCGTAAT	1080
TTTTTCGTTT	TTTCGTTTCG	TTTTTCGTTT	TTTTTCGGCG	GCGGCGGGGG	TCGGTACGGT	1140
ACGGTTTATA	GCGGCGGTTT	TTTTGCGTCG	AGTTTCGGAT	GTTGTTTTTG	GGGAGGCGGA	1200
GGTAGCGGTA	GCGGTAGCGG	TTCGGTTCGT	ACGGTTATTA	TCGTTGCGCG	TAGTAGGCGT	1260
TCGCGGGTCG	AGTTTTTTAG	TAGTCGGCGT	CGGCGTCGGC	GGCGGCGGGC	GCGGCGGTTA	1320
TGGTTTTTTG	CGTTTGTTTT	TTTCGTTTTT	GGCGTTTTGC	GGCGGTCGCG	GTTTCGATTTT	1380
CGTATTTTGC	GTTTGCGCGC	GGTTTAGTTA	GCGCGGCGCG	TTTTTTTCGC	GCGCGTTCGC	1440
GTTTATTTAT	AATCGCGTTC	GCGGAGTGTT	TTTCGTATCG	TTTTCGGTGC	GTGGGCGGGA	1500
GATATAAGTT	TAGTTAAGTT	TATATAGTTC	GGTTCGGGTG	GTTATGGAGA	GGCGTATTGT	1560
AGAGATATTT	GGACGCGTAT	TGTGATAGGC	GTTATTTTGA	TATACGTGTT	TTTTTATTTT	1620
TAAATATTAC	GGGTCGTAAG	CGTTTATATT	TATATTTATT	TGTAGGGATT	ACGTGGGAGA	1680
ATTTGAAGGG	TTGATTTTGT	TTCGGTGTAT	TCGTAGTAGT	AATGATGCGA	GAATTCGTTT	1740
TAATTTTCGT	ATTGATTGGA	TTAGATTGTA	AATTTTTTGG	CGTTTGGGAT	GGTTTAGTCG	1800
TGTTTAGTAT	TAAGTGTAGG	TATTTAAATT	TTGTGGAGT	AAGTGATTGG	ATGAATGAAT	1860
GTATTGGAAT	TGTTTATTGA	GTATTTGTAA	TAATTTTGAT	TTTATGGTTT	GGTATAAAGA	1920
ATTTTATTGT	AGTATGAAAT	TAAATGGTGT	TTGTGTGTTT	GTTTTGTTTT	GGTTTAAGGT	1980
TTAAAATAGG	TTAATGTATG	TTTTTTTTTT	TTTAACGTAG	AGGAGTTTTA	TTTTGTTTTT	2040
GTGTTAGTTT	TTATTATAGA	ATGCGTAATT	TTATTTTAAT	TTAAATGTTT	ATATTTTTTA	2100
GGGTAGGAAT	TGTTGTTTAG	AATATATAGT	TTGAAGTATA	TTGTGTGTTT	ATATAACGAT	2160
AGTTTTTTTG	GTTTGGTAAG	ATTTTTTATT	TTTTATTTGT	TTGTTATTTT	GTTTTATTAC	2220
GATTTTTTGT	AATTTTTTATT	TTTATTTTTT	TTGTTTTTTA	TTATTATTAA	AATTTTTTTA	2280
TTGTTTTTTG	TTTTTTTTTT	CGTTAGTTTT	TGGGGAGTTA	GTTGTTATGA	GTATTTAGTT	2340
TGGTTTTAGA	TGTTGAAAGG	GTTAGTGTAT	ATAGAAGATA	TATATTTAGT	GATGGGTAAG	2400
ATGTTGTAAT	TATATTGGGT	TATGTTGAAA	TTGTGAAGTT	TTAATTTTTT	TATGTGTAAG	2460
AGAAATTAAT	GTATTGTAAT	TTCGATGTTT	GTGTTGAGGA	TTTTTTTTGT	AAATTTAAGT	2520
TTTTGGTTTT	TGTTTT					2536

ANGABEN ZU SEQ ID-NO:11:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 504 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:11:

AGTAGTAGTA	GTAGCGCGTT	GAGTTCGGTT	ACGTAGAGCG	TTAGTTCGGT	GCGGTTTTGT	60
AGGTCGGCGC	GCGTGAGTAG	CGGTAGCGCG	AGTATCGAGG	CGTTTAGTAG	GAAGGTGGCG	120
AGGTTGAGAC	GCGTTTTAGA	GTTGCGCGGA	GTTGCGGTCG	CGTAGTTCGT	TATGGCGTCG	180
GGCGGCGGAT	TAGCGGGCGG	CGGCGTCGGT	TGTAGTTCGG	AGAAACGCGT	CGGTTGCGTT	240
TGCGTATTGT	GTCGTCGATG	TCGGTTCGGG	AAGGGTAGTT	GTTGCGTAGT	GCGCGTCGTT	300
TGTTAGTTTT	CGCGAGAATT	TCGGTTCGTT	TTGTTGTTGT	TCGGAATTTG	GCGGGAGCGT	360
TTCGTTTCGT	CGTTTTTTTT	CGTTTTTCGG	GGATATTCGG	GTTTTGAGTT	AGATTTTGTG	420
TCGGGCGGGG	GTCGGGAAGT	GGTGGGAGAA	GTCGTCGTTT	GCGTTTGTTT	AAATTTAGTT	480
TTAAATTTAG	GAGGTTTTTG	TTTT				504

ANGABEN ZU SEQ ID-NO:12:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 2036 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:12:



AGTAGTAGTA	GTTTTAATTT	GAGTTTAGCG	TATTCGAAGT	TTCGTTTACG	TTTTTTTTTT	60
TAGAGATTTT	TAAGTCGGAG	GAAAATAGCG	TAGTGTCCGT	TTTLAGTTTA	GTCGTTGTTT	120
TTTCGGTTTT	TAACGGTTTCG	GATAAGATGT	AGATGGAATT	ATCGTTTTTG	TTAGGTTTGA	180
GTTTTTATTA	GTTTTTTTAG	TAGTCGTCGT	CGTTTTAGGA	GTCGCGGTA	TCGGGCGCGT	240
CGTTGTCGTC	GTTTTTCGGT	AGTATTTGGT	TTACGGGTAT	TATTAACGCG	GTAGAGGATA	300
GTTTTTTTTA	GGGGATTATT	TTAGTTAACG	GGATTATGTT	TTTTTAGAAT	TTTTCGTATT	360
ATGTTAATTT	AGTTTTTCGGA	GGTATTTTTT	TTTCGTAGAT	CGGTTTGCG	TAGATTTAGT	420
ATTATTAGTA	GTCGTCGTCG	TTTGCGTTCG	CGTCGTAGTC	GGTATAGTTA	GCGTAGTTAT	480
TATAGGCGTA	GTTTTCGTAG	TAGCGTCGTT	TATTCGTTAG	TTTLAGTTAG	GCGTTTTACG	540
CGTAGAGGAG	CGTCGTCGCG	GCGTACGGTT	ATTAGTTTAT	TATGATTAGT	AAGTCGTTTT	600
CGTTTTTCGGC	GGTTGTAGTC	GTTGTTGTCG	TAGTCGTCGT	TTTCGTCGGT	TCGTTTAGTT	660
GGAATACGTA	TTAAAGCGTG	AATGTAGTTT	GGAGCGTATC	GTTTAATTTT	TGGGGCGGTT	720
TGTAGGCGGG	TCGGGATTTT	CGTCGGGCGG	TCGGTGTGGG	CGTGGGTGTG	GGTGTGCGGG	780
TGTTTTTTTC	GTTTAATTTT	ATTCGTCGT	TTAAAAAGTT	TTTTTTTAGT	AACGTGATCG	840
CGTCGTTTAA	GTTTTTTCGC	GCGGTTTTTT	TTATTTTTAA	GTTTTGGATG	GAGGATAACG	900
TTTTTCGGAT	CGATAATGGT	AATAATTTGT	TGTTATTTTA	GGTAATATTT	TGTTGTATTT	960
ATATATTTTT	TTTTTTTGTT	TCGTTTTTTT	TGTTATTTTT	TTGATTTTTA	ATTATTTATT	1020
CGTTTATATT	TAAAAAGGGT	AGTAATGTTT	AGTTTTTTTT	TTTTTTTTCG	AAGTTTTTAG	1080
TTTTTTAGGG	TTGTTTTATT	ATTAGAGGAT	GAGGTTGGGA	GAATATTGTG	ATTATTGGAG	1140
GAATCGTATT	CGTTTTTTTA	GGGTAGAAGA	AGTTTTTTTT	TTTTAGTTTA	TTTTTTTTTT	1200
TATGGTGTTT	GTATTTTTTA	TTTTGATTTT	TTTTGATAAG	TAAAGTTGTA	AGTGTGTGGT	1260
AAGGTGTCGG	TATAGTTTTA	GGATGAGTAG	GTTGAGATTT	TTATTATTTT	GAGTAGTTAT	1320
ATATTTTAGG	TTATGTAATT	TTGAGTTTAG	GGTGTCGTTT	GTAAGCGGTT	TATATTTAAA	1380
TTATTTTGAT	TTTATTTTTA	AAGAGGTTAA	TAATTTTTGA	GTGGTGTTAT	TTTAGACGGT	1440
TGCGTATGTT	TAGTTAAATT	AATGTGAATA	TATGTGTTTA	TGTTTCGTTA	AGGTGTTAGA	1500
ATTATTAATA	ATAATTAGTA	TATTGTTTTT	GTTGGGAAAA	TTATGAGTGT	GAGATTTTAA	1560
TAAATATTTA	TTATTGTGTA	AGGGAAGGAG	GTGGAAGAGT	GGAAATTTTA	AGGGTAATTT	1620
TTTTTATGTT	ATTTAGAAGA	AGGATTTTTT	TTTTTTTTTT	CGGAGGTAAG	AGATAATAGG	1680
ACGTGATTTT	TAGAGTTATA	TTGTAATGGA	GTTTATTTGT	TAGTTAGTAT	TGAAATATTT	1740
AAAGTTGTAG	TGTTTTTTAT	TAGTTGGGTT	AAGAAAATAT	AGATTAATTA	TATTATTTGA	1800
AATTGGGTAA	AGTGAAAGTT	TTGTAATTTT	AATGAAAAAT	GAGTTTTGTG	TTTTGGTGTT	1860
AAATGTTAAT	GAAAAATTTA	GGATGTGTTT	TAGATAAAAA	GGTTGAAATG	AAGAGTGTTT	1920
TTATTTTTGT	TTTTGTTTTT	TGATTTAGTT	TATAGTTATA	GTAATTGGTA	GTGTTTTGGG	1980
TATTTTGGGT	TGGAAAGGCG	TAGTGCGGGT	GTGTGTTTAT	AGGCGGTTTT	TGTTTT	2036

ANGABEN ZU SEQ ID-NO:13:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 452 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:13:

AGTAGTAGTA	GTAATTTTTT	TGTATTTGTA	GTAGAAAATT	TATTTTAGAG	ATTGGGTATG	60
TGTGTTAAGG	AGATGGTGAT	TTTTTTTGGA	TATTTTAGAG	ATTTTTTTTT	TTGTTATGAT	120
TTTATTCGG	GGGATATTTA	TATAATGTTA	TTTTTTTGA	TTAGTTTGT	TTATATTAAG	180
ATTGTTATTT	TTTLAGTTGT	TTGTGTGTTT	TTAGATTATG	ATATTTATTT	AGTTGTTTAG	240
TTTAGGAAAG	ATTTTTGGGT	GTATAAATTA	GTTTAGAATA	GTTGGTAGTA	ATTTGTAGAT	300
CGTTTGTTGT	GTCGTGAGGG	TATGGATGTG	GTATATTTTA	ATGGTTATAT	TTATATTTTG	360
GGGGACGAG	ATTTTATTAT	TGGAGTTAAG	TTGAAGGAAG	TGGAATGTTA	TAGTGTTTAG	420
AGAAATTAGT	GGGTATTGGT	GGTTTTTGTT	TT			452

ANGABEN ZU SEQ ID-NO:14:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 513 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:14:

AGTAGTAGTA	GTAGCGGAGA	CGGAGGCGGC	GGCGGAGGCG	GCGGCGGCGA	TTT TAGCGCG	60
GGGCGAGCGG	CGGGGTTTAT	GGTCGGATTC	GGCGAGTTGG	TTTAATAGTG	GGGT TAGGGG	120
TTTCGGGCGC	GGGGGGGATC	GGAGGGAGCG	AGGTCGTTGT	CGGACGGGGC	GGGGCGTCGG	180
GAGGGGCGGG	GTCGCGGTCG	GACGGGGCGG	GGTATTAGGA	GGAACGGAGT	GGGCGTTTGG	240
CGGTTTTTCGC	GTTTAGATTG	GGTCGCGGGC	GTTTTTTTGGT	GGTCGCGGAG	AGTTTAGGTG	300
TTCGCGGTTG	AGGGAGTTGG	AGAGGGGTAA	AATCGGGGTC	GTAGCGGGAA	GGCGGAAAGT	360
TAGAGAGAGG	AGGTGTTTCG	GTAGCGGACG	CGTTAGCGGG	GTAGATAGGA	AAATAGTTGG	420
AAAGGGCGAT	TTGGGGGAGT	AGAGATACGA	TTAGAGTTTG	AGAAGGGTAG	GAGTAAGGGA	480
TGTTTAAGGG	GTTTGGTGGT	AGGTTTTTGT	TTT			513

ANGABEN ZU SEQ ID-NO:15:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 980 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:15:

AGTAGTAGTA	GTTTTTTCGG	TTATAAGGAT	TTTTTCGAGTT	TCGTTTCGTCG	GTCGCGGATT	60
CGGTTTTTTT	TTTTTCGGTC	GTTAGGGGGC	GGGTTCGGAT	TATAGGATTG	GAGTTGGGCG	120
GAGATTTACG	TTCGGAGCGG	TTGTGAATTG	GTAGGCGGTG	GGCGCGGTTT	TTGTGTCGTG	180
TTTCGGGTAT	TTAGTTTTTT	AACGGGGTTT	CGGAGTCGAA	GATAGTTTTA	GGGTTTAGGG	240
AGCGCGGGCG	GTTTTTGGGC	GGCGTTAGAT	TGCGGTGAGT	TGGTCGGCGT	GGGTATTAA	300
TTTAATGTAG	TTTAGGGCGG	CGGTACGAGA	TAGAATAACG	GCGAATAGGA	GTAGGGAAAG	360
CGTTTTTCGAT	AGGTTAGGTT	TAGGGATTTG	CGGGGAGAGG	GCGAGGTAA	TATTCGGTAT	420
GGGTTTTTGA	TTGGTTTTTG	GGATTCGTTT	CGTTTACGTT	TATAGGTGGG	TTCGTATTTT	480
TTTTTGC GTT	TCGTTTTTCGT	TTTAATAGTT	TATAGTTGTT	GTTAGTTTAT	TCGTACGTTT	540
CGAATTTTCGT	TCGAATTCGT	TATTGGTTGT	TTTTTAGCGG	TTTGTGTTGA	TTGGTTGTTC	600
GAAGATTCGT	TTTTTTGTCG	TGGGTTTAGT	TTCGTAAATG	CGTAGTTAAG	CGGGTG GTAA	660
GGGGCGGGTG	GAGCGCGGGG	CGCGACGGCG	GAGGGGGGCG	TGGGTAGTCG	GACGTATTTT	720
GGTAGGGAGT	AGTAGGTGGC	GGCGGTGTAT	GGGGTTTGGT	TTTATTAGCG	GGTATTGGTT	780
TATAGTTACG	GTCGGGGGGT	TATTTAGTTG	GAGAGAGAAG	GGATAGGTGA	TTCGATCGGA	840
GTTTAGTTTA	GTTTTTAGCG	GTGGGGCGAG	AGATAGCGAG	GGGAATCGAG	GTTGGGGAGG	900
TTATTTAGGG	AGATTTTCGA	GGGAATTTGG	TGAGGTTTGA	ACGGAGGGAG	ATTTGGGGTT	960
GAATAAAGGG	TTTTTGTTTT					980

ANGABEN ZU SEQ ID-NO:16:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 223 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:16:

AGTAGTAGTA	GTAATATTAT	TTTATTAAAA	AAATATTAGC	GTAGTTTTTA	ATAGAGTTTT	60
ATTTTTTAAA	GTGAAGTTAA	AAATAGGTTT	TATGTTTAAA	GTTTTTTTTT	GTAGTAGGCG	120
ATACGCGTAG	ATGGAGAAAA	TATTTTTATT	GTTGGGAAAG	GGAGGCGTGT	TAACGGGGAC	180



GAAGGATATT TTATTATTTG GTAAAGTTAT TGGTTTTTGT TTT

223

ANGABEN ZU SEQ ID-NO:17:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 1145 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:17:

AGTAGTAGTA	GTTTTGGGTT	TCGTTTAGTC	GTAAGATGAG	TTAGTTGGGT	TTGGTGGGGG	60
TTGTATTTAG	GTTACGTTTT	TTTAGATTCG	TTTTTTTTTA	TTAGATTACG	TTTATCGTTT	120
TTTAGTTCGT	TTTTGGTTTA	GGTTTTTTTT	TGTTTTTTTC	GTTTTGAATT	TTGTTTCGAG	180
TTTCGTTTAT	TTTTTTCGGT	TTTTTTTAGT	TTTGTTTTTC	GGATTTTGGT	CGAGTATTTT	240
AAGTTTTTTT	TTTGGCGATT	AGGTTTTTTT	ATTATTTAGG	GTTTCGTTTT	GTTTGTTTCGT	300
TTGGGGCGGT	TTGTAGTTTT	ATCGTCGAGG	AAGAGGAAGG	CGTTCGTGGT	TTGGTTTCGT	360
AGCGATTTCG	TTTCGTCGTA	TCGAATTCGT	AGTTTTATTA	TTATGTTTAT	TACGTCGATA	420
TTTATCGTTA	GTATTATGTA	CGTTATAGTT	ATCGTTAGCG	TGGTTGGGGA	GAGCGGGCGT	480
ATTAGTTTTT	GCGGGCGTCG	TTTTAGTTAT	TACGGGTTTC	GTCGTAATCG	GAAGTTTTTT	540
CGTTTTTTGT	TTTTAATTCG	GTGTGTGGTT	AGGAGTCGAG	TTTTATCGAT	CGGGAGTTAG	600
GCGTTTGGAT	TGAGATTTGG	AGGTTATATT	AGGATTTGCG	GTTATTGTCG	TTATTAGGAC	660
GGGGAGGGGG	TCGGGGATAA	GTTTATAAGT	CGAGGGGCGG	GGTTCGTGGC	GGTTTCGGTC	720
GATAAGGGAT	AGTTGTAATT	GAGTTCGGGG	TGGAAGGGGT	GGGCGTTTCG	TTATGGGTTT	780
AGTTGTGTGT	TTCGAGTTTT	GGGGGTTTGA	AATGTTTTTT	GATATTTGGG	CGGGCGGAGT	840
TTTATTTTGG	TAGGGGATGT	TGGAGTAGAT	GTCGTGGTTG	TATTTTTGTT	ATAGGTTATT	900
GTGGTTTTTT	TGTTGGCGGG	TTAGTAGATT	GGTGGTCGTG	GAGAGTATTA	TGAGGACGTT	960
GGTACGAGG	TTGAGTAGAA	TGTTTTTATT	TTGGAGGTTT	CGTTTTATTT	TTATGTTATT	1020
GAGGTTGTAG	TCGGGGGTTA	TAAGGGTAGG	ATGGGTTTAG	GTTTTTTTTA	TTTCGAGGTT	1080
TTTAGTTGAG	GAGTTCGTTT	TTTTTATATT	TTTTTTTTTG	TATTTTTGTT	TGAGGTTTTT	1140
GTTTT						1145

ANGABEN ZU SEQ ID-NO:18:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 633 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:18:

AGTAGTAGTA	GTATAGTTCG	GCGTTGGTTA	GCGGTTTAGA	GTCGGGGTAG	GGAGGGTTTT	60
CGTGCGGACG	GGGGTTAAGA	GGTCGTTTTT	AGTTCGGGGT	AGTTCGCGTA	TTTCGAGTTT	120
GAGTTCGGAG	TTTGAGAGGG	TGGGGAGGTT	TCGCGTTTAG	GGTCGGATCG	GAGGGAGAGG	180
GAGAGCGGTG	GTTTTTTTTT	CGGTTTTTCG	CGTTAGTCGA	TCGGGGCGTT	GGGCGGGCGT	240
CGCGGGAGTC	GTAAGTTTTT	TCGGGGGGCG	GATTCGGTTT	CGGCGGCGGT	ATTTAGTTTT	300
TGTTTCGATA	GGTGCGCGTC	GCGCGAAGGA	ATGTAGTCGG	TTTGATTTAT	TAGCGTTTTT	360
TTTTTATTTG	CGCGTTCGTT	TTATAAAGCG	TTGTTCGTTT	CGTTTTTATT	TTTTTAATTT	420
TCGCGTTCGT	TTTCGGATAG	TTTTTGTTTC	TTCGCGCGTT	GTAAGTTTTT	TTTTTAGCGG	480
TAGTTTAGGC	GCGGAGGGAG	CGAGTTCGTT	TCGAGGTAGG	TTTAGGACGG	GCGTATAGTA	540
GTAAGTCGAG	TTGGTCGGGA	GAGGGTGAGT	GTTTCGTTTA	GTTGGGTAGT	CGCGTAGGGG	600
AGTTGGTCGG	GTAAGGTCGG	TGGTTTTTGT	TTT			633

ANGABEN ZU SEQ ID-NO:19:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:19:

AGTAGTAGTA GT

12

ANGABEN ZU SEQ ID-NO:20:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 12 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:20:

ACAAAAACTA AA

12

ANGABEN ZU SEQ ID-NO:21:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 74 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:21:

AGTAGTAGTA GTTTCGGTAG TTTAGTTTAT GCGGCGGTG GCGGCGGTAG TAGGTTTGAG  
TTTAGTTTT TGTT

60

74

ANGABEN ZU SEQ ID-NO:22:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 103 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:22:

AGTAGTAGTA GTAGCGGTAG CGGTAATAGG GCGGTTGAGA ATTCGGCGGC GCGGTTTTTT  
TTCGTTTTTT TTTTTTTCGT TTCGTCGATT TTTAGTTTTT GTT

60

103

ANGABEN ZU SEQ ID-NO:23:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 559 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:23:

AGTAGTAGTA	GTAAGAAGGA	AAAAAATAAA	TTTTTTAAGT	ATTTAGTGTA	GTTTTTTTTG	60
GGTTTTTTAG	TTTTATTTTCG	GTATAGTTAG	GGTTTTTAAT	TTATTTTGT	TTTGGGTTTA	120
GGGAAAATAA	AAAGTTAAGT	TTATTCGTAT	TTAGTTAAAA	GTTGAAAATA	AGTGGTCGGA	180
TTAATTTTTG	GTTTTAAATA	AATTATTTGA	TGGTGTTAGA	AGGTAATAAA	TTTTATTTTT	240
TTTTTTAGAT	AAGGGTTTAT	TTTACGAGAT	TATTAAGTAT	TGTTTCGTGA	GATTATTGAA	300
ATAATAGTAT	TTTTTTTTTT	ATCGAGGATT	ATAGTATGTT	TTTGATTTAA	AACGTTTTAA	360
GTTTTTGAAG	AGATATTG	GAGTTTTTAG	AAATTAAAGG	AAGGATTGGA	AATGTTAGGG	420
GGTGAGGGGC	GTTAAATAAT	TTTGATTTTT	ATGGTTGTGA	TTGATGTTGT	TATTAGATTA	480
TTTTTATATT	TAATAATATT	GAATTATTTA	AACGGATTAG	TATAAATTAA	ATTATGATAG	540
TATATTTTTA	GTTTTTGT					559

ANGABEN ZU SEQ ID-NO:24:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1695 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:24:

AGTAGTAGTA	GTTTTGGAGA	GTAAGTGGGT	TTTATGTAAA	GTGTTATTTG	TTATATATTG	60
TTATTTTTAN	AGGATTTTAG	AATATTTTCG	AATTCGAAAA	TAGAGTGAGT	GTGGAGGGGA	120
GGGGGGTGTT	TACGTGGAGG	AGGGTCGCGA	GAAGGAGAGT	GCGTCGGGTT	GATGGTTAGC	180
GGTGTTTCGG	GGGTCGATGT	CGGGCGGGAT	CGGCGGCGCG	GGGGTGGGGC	GACGGCGGGG	240
CGGCGGTCGA	GTAGTAGAGG	GGCGTTTCGT	AGAGGTCGGG	GGGGGCGCGG	GGTTTCGCGT	300
CGTTTTGTTA	GTAGTCGTTT	CGTCGTTTTT	TTCGGTTCGT	TCGTAATCGT	CGCGGGATTC	360
GTCGGTTCGG	GTTTTTGGA	TTTTTTACGG	GGGATAGTCG	ATCGGGGGCG	TCGGGGTTCG	420
TTTTCGTATT	CGATTCGGAC	GGCGGTAGTA	GGGGGAGGGA	TGGTGTATCG	AGGAAGTTAA	480
GGTTTTATTA	TTTTGTTCGG	CGGGTCGGCG	GTTTCGTTGGT	TCGTTGGTTC	GGAGAAGTGT	540
TGCGTTTAGG	TTGGTTCGTA	GGAAACGGCG	GCGGCGGTTT	ANTTNTANTT	TTNNNNNTNT	600
TNNTTGNNAN	TNTTTTTTTA	TAGGGGAGGG	GTAGGCGGTT	CGCGGGTTTC	GCGTCGGTTCG	660
GGCGATTGGG	TACGCGAGGG	AGCGGGTAGG	GTAGGGGGAA	ATAAATTAAG	GTCGAGATTT	720
AGAGTCGGAG	AGCGCGGGAG	GAGTAGCGGC	GAGAGGTAGG	AGAGGTAGAG	AAGAAGAAAG	780
GAGGGAGAGA	GGGGGCGAAG	ACGGTCGGGA	AATTCGGGGT	TGTAGTTTCG	CGTCGCGTCG	840
GAGTCGTGAC	GGATTTATTG	TAGCGGTCGC	GTTGATTTCGT	AATTTTAAGG	TACGAAAAAA	900
GGGGAGGGTA	GAGAGGGAGG	GGAAAGCGGA	GTGTTGTAGC	GTCGGGGCGG	GGGCGGGTTT	960
TCGCGAGCGT	CGTATATTCG	GGAATTTGTN	GTTTCGTTGC	GGGCGAGTGT	CGCGTGTTTT	1020
GGCGAGTTTT	GGTTGGGGAA	NTTTGAGCGC	GCGGATTACG	TTCGTTTTTT	AGTCGGTTCG	1080
TTTTTGCGTG	TAGGGTGGGG	GGAGGTTAGG	GGCGGGGGCG	CGGACGTCGG	TGACGTCGCG	1140
TGGCGGAGTT	TTTTCGGTTA	TGTGGTTGGA	GGCGGGCGGG	GAAGAAGGTA	AGGTTGGGAG	1200
GGGAAGGAGG	AAATGCGAGG	GTTTTTCGGC	GGGAGGAAGC	GCGAGGGGTG	CGCGAGTGTT	1260
AGCGAAAGGG	GGATAGGGGT	TTTGGAGTGG	AAGTTTTGCG	GGAGAGGAGG	AGTAGAGGAG	1320
TAGTTTTAGG	GTGTTAGAAT	TATTCGGATG	TCGTATTAAA	AAATAGGAAA	AATTTAAGTT	1380
TTGCGAGAGG	TAGTTAAAGG	TATTTGTGTA	TTCGTGTATC	GATTTGGATT	TATTATAGAA	1440
TTGTACGGTT	TATTAGGATT	GTTGTTATTT	CGGTGTAGTA	TTCGTAGGTT	TATTTGTAAT	1500
ATGAATTGGT	AGAGTTAGGA	ATTAGGTTGT	AAAGATAGAG	AAGCGAGTTA	AAGTTGGTCG	1560
TAGTTCGAGG	CGTGGGGAGA	ATTTGGGTAA	ACGAGGAGAA	GGGATATTTT	TTATTCGTAG	1620
AAAGATTTTT	TATTTAGTTT	TGTATTTTTA	TAAATCGTTT	AGATTTTTTT	TTGGCGGAGT	1680
TTATTAGTTT	TTGTT					1695

ANGABEN ZU SEQ ID-NO:25:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 722 Basen

ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:25:

AGTAGTAGTA	GTTTCGTTAC	GAATACGATT	AGTTTATATT	GGTTGTTTAG	TATTTTCGGTT	60
AGTTTTTGTG	GGACGTTGGC	GCGTATATCG	TTTCGTCGGG	GGGTCGGGGA	TTTCGGGGAT	120
TTCGGGGTCG	AGGACGAGGG	AGGCGAGTAG	GTCGTTGGTT	TCGACGATTT	TGCGATCGAG	180
GCGTTTGTAG	TTGTTTCGGG	CGAGGCGGGC	GATTTTAGGT	TTTCGAGTAT	TTTCGAGATA	240
TTTCGCGTTAA	GTTTGGTTTC	GGGGGTGGTT	GTCGTCGGGT	TTTCGTTGCG	TTTTTGGTTT	300
TTGTTTTTGT	TTTCGACGTTT	TTGTTAATTA	GTATTTTTTTT	TTTTAATTTT	TTTTTCGATTT	360
TTATTTTACG	TTTTTGTAA	TTTGTTTTTT	TTTTTGTAT	TTTTTCGACGT	TCGTTTTTTT	420
TTTTTTTGT	TTTTTCGTTT	TCGTTAAGGT	ATTATTTTGT	TTATTTATTT	AGCGTTTTAT	480
TTTGTGATT	TGGGATTTTA	CGAGTTTTTT	TGTTTCGTTGT	TTTTTATTGG	GTAACGTTTCG	540
GGGTAGTTAT	TTGTTTTTTT	CGGGATTTAC	GCGGATAGTT	TTTCGTTTTT	GATTTTTTGGG	600
ATATTGGTTA	GTTTTGTTCG	GATATTAGCG	CGTNTTTTCG	TATTTTCGTT	CGCGCGTTTCG	660
GTTTTTCGTT	GTCGTTTGTA	GTTTTTATTT	TTGAGTCGAC	GTTTCGTTTAT	TTAGTTTTTTG	720
TT						722

ANGABEN ZU SEQ ID-NO:26:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 517 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:26:

AGTAGTAGTA	GTTCGAATTC	GCGCGCGTAG	CGGCGTATGG	TTAGGTTTAG	CGAACGTAGT	60
TTTAGCGAGT	GGCGCGTTAG	GCGTATTACG	TAGAGTACGC	GTAGCGTACG	TAGTAGTCGT	120
AGTATTAGTT	TCGCGCGTTT	TAGGAGTTTG	GTTTCGTTTCG	GGTTTGTCGT	TAGTTTTAGT	180
AGTAGCGATA	CGTAGAACGG	TAGGAGCGTT	AGGATGTTAA	TGATGTTGAG	TGGCGCGCGT	240
AGGAAGGCGT	ATTTGTTTTT	GGTTTGTAGG	GAGCGTAGTA	GGAATTCGAA	GGAGAATTAG	300
GTTACGTATA	CGGTTTTTTAG	TACGAATAGG	TTGCGGTATT	TGGGGGAGTA	TTTCGTTTTGC	360
GGGGAGAGGG	GATACGGGGT	TGGGGGCGTA	GGTTTTTTTG	AGGGTCGGGG	GCGTTTTTGT	420
TTTTTTTTTT	TAGGATTAGA	GCGGTTTTTT	AGGTTATTTT	TTTCGTTTTT	ATTTACGGAA	480
GCGGGTGTAG	GGGATATTGA	TAGGTTTAGT	TTTTGT			517

ANGABEN ZU SEQ ID-NO:27:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 1078 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:27:

AGTAGTAGTA	GTATTAGTAA	TAGTAGTACG	AAAAGTAAAA	TTGTAATTTA	AACGGTTTTT	60
AGGGTTTAAG	TAGGTTTCGAC	GAAGATTCGT	TTATTCGGTC	GTTAGAAAAT	TGGGAGTTTC	120
GTTTCGTTTTT	TCGGTATTGA	AACGCGATCG	GTTTTGTTTG	GTATCGTATT	TTTTTTTTGA	180
TTTTATTTAT	ATTACGACGA	CGGACGTTTC	AGAACGTTGT	ATCGCGTTTC	GTAGGAAGTG	240
TTTTTTTGGG	CGGAAGTTTT	TGAGCGTGAT	ATAGCGGAAG	TGTTTTTTTT	TTTCGTTTTT	300
TTGGTTTCGG	TCGTAGAAGC	GAGATGGTGA	GTTGTGATTG	TGGTGTGTTG	GAATCGCGTT	360

TTATTTTCGT	TTTTTGTTT	TTTTTGTTT	TTGTGTTTG	GGGGTTGGTA	AGATTTTCGGA	420
TAAGGGGAAT	TGGTGGGTTG	GAAAGAGGTA	TGCGGTGGTT	TTTAAGAGTT	AGAAGAATGA	480
TTGTTAATTG	GTGTTTGGG	GATTTATTTT	GTCGTAATTG	TGGTGTTAGA	GTCGTATTGT	540
GTTTTTTGTT	TCGGTTTAAT	TTTTGGAGAT	TTTTTACGGT	TTTAGTTTTG	GTTGGGAGTC	600
GAGGGAAGGA	GTTTGGGAAT	GTTTGGTTTT	GTGTAATAAT	GAAATAATTT	ATTGGTGATG	660
TTTTTTGGTC	GGAGTTTGTA	AAGATAAGGT	GTATTTTAGA	ATATTGTAAT	TTTTGCGGAG	720
GGTTTTAGGT	AACGTGAAAT	GCGGGTAGTG	GTTTTTGATT	TGGTATTCGT	GGAAAGAGGT	780
TTTTTTTCGTA	GTTTCGATTT	TTACGATTGT	TTTTAAATTT	TATTAGTAAT	TGGTTTTTCG	840
GAGAATTCGA	GTAAATTTAG	AAGTTGTTAG	GTTTTAGAAT	TATTTATTTT	TTTAATGTGT	900
AGACGAAGGG	AACGTTATCG	TTTGGAAAGC	GTCGTAATAA	GACGTATACG	TTGTGTCGTC	960
GTTGTGGTTT	TAAGGTTTAT	TATTTTTAGA	AGTCGATTTG	TGGTAAATGT	GGTTATTTTG	1020
TTAAGCGTAA	GAGAAAGTGT	AAGTAATATT	TTTTAGGTTA	ATTGTGTTAG	TTTTTGTT	1078

ANGABEN ZU SEQ ID-NO:28:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 2949 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:28:

AGTAGTAGTA	GTCGTAGGAG	TAGCGTTTTG	GGGAGGGGGG	TTCGTTTTTC	GGGGTGCGTT	60
TCGTGGTTAT	TTTTTTTAAN	GGGGGTCGTG	GTCGTTTTTT	AGTTTCGGCG	CGTTGGGTTT	120
GTTTTAGCGG	GGTTCGTCGG	GCGGGGCGGG	GATTTGTATT	CGGGGCGTTT	TTTCGGGACG	180
CGCGGGTTTT	CGTTTTTTTT	TTTGTTTTCG	GGCGAGTTCG	GATTTTGTTT	GGTCGCGGGG	240
TTTCGTTTAT	TTGTTAAGTG	AAGGTTTTAC	GGGAGATAAT	AAAGAAAGAA	GTTGTTTTTT	300
TTTTTTTTTAG	TTAAATTACG	GATGATTAGT	AATCGGTTTC	TTTTTTTTTT	TGTTTCGTTTT	360
TTTTTTTTTTT	TTTTTTTTTT	TTTTAGGAAT	ATATTTTATT	CGTTTTTTTT	TAGAGTAGTA	420
AAATTTTTTC	GTTTTTGAAG	GCGTGGTGTT	GTTGGTTGTT	CGCGAGTGCG	GAGAGGGGAG	480
GTTGGTTTTT	GTTTTTTTTT	TTTGGTTTTT	TCGTTTTAAG	TTTTGTTTTG	TTTTAGGAAA	540
ATTTTTTGCG	ATGGTGTATT	TTTTTGAAGT	TGTTATTGTC	GGAGTTTTTG	GTTGTTTTTG	600
AAATAAGAGT	TTTTTTTTTAT	TTTTTTCGTA	ATAGTTATAG	TTTGATGCGA	TATTTTTTTT	660
TTTGGGTAGA	AATTAGGAAA	TAAATATAAT	AAGGAAAAAA	TTGAATAGGA	AAAAAAATGA	720
TGTCGAATCG	TTTTGTTATT	TTGTTGTTGA	CGGTTGGGGT	TGGTGTTTTA	TTTTTTTTTT	780
AGTAGTTGTT	TGGAGAAGGA	GGAGGAAGAA	GAGTGTATTG	GTGGAGGTGG	AGGTGTGTGT	840
GTGTATATAG	GGGATCGATT	TTTGATATAT	AATAATTTTA	GTGGGTACGT	TGTACGGTTC	900
GTAGAGTGTA	TCGACGGCGG	GCGTTGGCGT	TGGTCGTGTT	GGAGTAGAGT	TGTAGTAGGA	960
GTTTCGTTTTG	CGTTTTCGTT	CGCGTTTTTT	TTTTTTTTTT	TATTTTTTTT	TTTTTTTTTT	1020
TTTTTTTTTTT	TTTTTTTTTCG	GGCGTTCGTT	TTTCGTTTCGA	TTTGATAGTT	CGTTTTTTGT	1080
TTTTTTTTTTT	TAGAGCGGGG	ATTGTCGAAT	GTTTATTTTT	CGTCGCGAGC	GCGTATTTAT	1140
TATTATATTT	TTTTAAATAT	AAGGAAGTCG	AGTATTAGGG	TTTTCGTTGA	TTGGTTATCG	1200
TTGGGTGATT	GGTAGGTTTC	GAATGATTGG	TTTAGGACGC	GAGGTTTTTT	TTTCGGTTCG	1260
TTTCGATTGGT	TGTTTTTTTAA	TTTAGGTAGA	GCGTCGTAGA	AGTTGGAAAT	TTGTCGTTTT	1320
TTAGTTTTTTT	TTTTTTTTTAT	TATTTTTTTT	TTTTCGTTTT	TTTTTCGTCG	TTTTTTTTCG	1380
TATATTTTAA	TTTTTAGTGT	TCGGTTTAGA	CGTTGGCGTT	TTTTCGGCGG	TTTTGGCGTT	1440
CGTAATAGGT	TTGGGCGGGG	GGAAGAAAGG	GGAGATAAAA	GGGAGGGAGG	GACGAGAGGG	1500
GGGAAGAGAA	TTAGAAGGAA	AACGAAGGGG	GAAATATGAA	AAATAGTAAT	TTGTTTATTT	1560
TATTTAGTAC	GCGTCGGGTT	GTTTATTTTT	TTATTTTCGT	TTCGTTAGAG	ATTTGTAAAG	1620
CGCGCGGTAC	GGATGTATTA	TTTAAGTTAA	TATTTATAGA	TAGATGTGTT	TTTAGTGGTT	1680
TTAAAGTTTT	TTTTTTTTTT	TATTAGTCGG	TTTTTTTTTT	TGATCGGCGG	GGGTGGGAAG	1740
CGGGGGTTTC	TATCGTTAAG	AGTTTTGGTC	GTGGTTACGT	TTTTTGGAAA	TTTCGTTATCG	1800
TAATTTTTTC	TAGTCGTTGC	GCGTCGGCGG	TAGTTTTTTT	TGTTTTAAGTT	TTTCGGATGTT	1860
ATTATTTGGG	AAATTTGTAG	TTTAGGANNT	TCGGTGTTTC	CGCGTGCGAG	ACGGATCGTT	1920
TCGTGTCGTC	GTTTTTTGGT	TTTTTTATTT	TTTTTTTTTT	TTTTTCGTTT	TTTTTTTTTT	1980
TTTTTTTTTTT	TTTTTTTTTTG	GGATTCGTGT	AAGTAGCGTG	CGTGTTTTTG	TGAGTTTGGA	2040
GGGTGTCGGT	TAGCGGAGTC	GTGCGATGTA	TTTGAAAAGT	AATTAGTTTC	ATTTTTTTTT	2100
ATAGAGTTAT	CGTAAGTGTA	TTTCGATTTA	ATGGTTTTAG	AAATTTTTGT	GGTATTAGGA	2160
TTTCGGTCGAA	AGAACGGGGA	TCGGTTATTC	GCGTTTTTTT	TTGTTTTATC	GTTTTTTCGT	2220



GTGTGTGTGT	GTGTGTGTGT	GTGTGTGTGT	GTGTGATTAG	CGGTGGGGGG	GTTTGTTTCGT	2280
TTTTTATTTT	TAGGTATGGT	GGTGTTTTTT	TTTTTTTTTT	TTTTGTTTTA	GATGGATTTG	2340
CGTGGTGGAT	GGGGTTGGCG	GCGATAAATG	TTTTTTTAGT	TTTAATTTAG	TTGAAAGAGT	2400
TAAGGGGGAT	GGGAGGGGGA	GTGTATTCGG	TAGGCGAGAG	AAGCGGGGGT	GGGGTGGGGT	2460
GGGGTGGGGG	GACGCGGTTA	AGCGGAAACG	TTGTATAGAG	GAATTTTAGC	GAATTAAGAA	2520
AAAAGAGAAA	GTGGTAACGA	AAATAAGGGT	AAATTGAGTT	TTTTTCGGGG	ATTTTAAATG	2580
AATTAATTTA	ATTCGGATAT	TTAATAAATA	TATGGTTTTT	AATGAGCGTG	CGTGTGCGTG	2640
TATTTTCGTAT	TTTLAGTTGC	GGGTGCGTTC	GTTTGCGTTC	GTCGTTTTTA	GTTAGAGTTT	2700
GTATTTGGTA	GTTTCGAGTT	CGAGCGGTAG	TTTAGGACGT	AGTCGAGGAG	CGTCGTCGGT	2760
GCGTTTCGAT	TAAATGTGA	ACGGGTTTGT	TTTTTTTTTT	CGTTGTATCG	TGTTTTTGTC	2820
GAGCGTAGTT	AAGTGAATTT	AGTGAAAGTA	GGAGTTTTTT	TGTTTAGTCG	TAGTTAGGTA	2880
GTTTTCGCGT	GATTTTAAGA	TTAATATTAG	ACGCGTAGAA	TTTATGTAGG	TTTTTGTTTA	2940
GTTTTTGTT						2949

ANGABEN ZU SEQ ID-NO:29:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 117 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:29:

AGTAGTAGTA	GTTTATTTTG	GGTTGATATA	AAATTGAAAA	ATTTATTTGT	TTTTTTATTT	60
TTTTGGGTTG	GATTCGGTGT	ATCGGTTGAT	ATATTTTTTT	GGTTATTAGT	TTTTGTT	117

ANGABEN ZU SEQ ID-NO:30:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 639 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:30:

AGTAGTAGTA	GTATTATGGT	TATTTTCGACG	GTCGCGTTCG	TTATTTTTTT	GCGGCGGTTT	60
AGTCGCGACG	TTGTTAGGGT	TAGTAAATTT	TTTTTATTTT	TGGGCGTAAT	GGTTGTCGGG	120
GTCGGGGTTT	TTTCGCGGTT	AGGAGGCGGT	GATTCGTTTT	GGGTTTGGGT	TTGTATTTTT	180
TCGACGGTTG	TTTTTCGTTT	TTTTTATTTT	GTTTTTGCGG	GTTTTAGTCG	CGGCGGTGTC	240
GTTTTTTAAG	TCGTTTCGTTA	AGGGGAGGTT	TTTCGTGGGT	TACGTTAGGG	GTAGTTTTTCG	300
ACGGTTTAGA	GTTAGTGGTT	TTTLAGTATT	TTTTCGTTTA	GTTTTAACGA	TTTTGGGTAT	360
TTGAGATTCG	CGGTTTTTTCG	GACGGTTGGT	TTTTTAGGGA	TTTGAGATGT	TTGTTTTTTA	420
GATTGTTGTT	TTTTTAGGGA	TGGCGCGGTG	TTTGGGTTTT	AGATTGTTTA	GATAGATTAT	480
TTTTTGATGG	AGAGGGGATT	GTTTTTCGCG	TTTCGGACGT	TTCCGGTTTT	GAGTTGCGGG	540
TGTTGTTTAT	CGGGCGCGAT	TTTTTAGTAG	GTTTTGCGTT	TTGTTTTTTG	GTAAGTATCG	600
ATTTATTTTG	TTATTTTTGT	CGCGGTTTTA	GTTTTTGTT			639

ANGABEN ZU SEQ ID-NO:31:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 304 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA



## SEQUENZBESCHREIBUNG: SEQ ID-NO:31:

AGTAGTAGTA	GTATAGTTCG	GCGTTGGTTA	GCGGTTTAGA	GTCGGGGTAG	GGAGGGTTTT	60
CGTGCGGACG	GGGGTTAAGA	GGTCGTTTTT	AGTTCGGGGT	AGTTCGCGTA	TTCGAGTTTG	120
GAGTTCGGAG	TTTGGAGGGG	TGGGGAGGTT	TCGCGTTTAG	GGTCGGATCG	GAGGGAGAGG	180
GAGAGCGGTG	GTTTTTTTTT	CGGTTTTTCG	CGTTAGTCGA	TCGGGGCGTT	GGGCGGGCGT	240
CGCGGGAGTC	GTAGTTTTTT	TCGGGGGGCG	GATTCGGTTT	CGGCGGCGGT	ATTTAGTTTT	300
TGTT						304